Free Will: Agent-Causation Revisited

Thad Botham

Abstract

This book maintains that the best theory of free will is an agent-causal view. The signature of agent-causation is that an agent acts freely only if there is an event having a cause, where this cause is not some other event but rather the agent of the act in question. With Timothy O'Connor, I reason that nothing can efficiently cause, though something may contribute causally to, someone's directly free action. I propose a theory of causal contribution in the spirit and on the shoulders of Roderick Chisholm.

I attempt to get very clear about several pre-theoretic features of free action. Although widely discussed at the pre-theoretic level, scholars rarely use the notion of ultimate origination when constructing rigorous arguments for, against, or in defense of particular theories of free will. Intuitively, an agent acts freely only if she is self-determining in that she is an ultimate source or underived originator of change. I argue only agent-causation satisfies a precise formulation of the origination condition. Accordingly, I fault Carl Ginet and Stewart Goetz's simple indeterminist theories as well as Robert Kane's causal indeterminism.

I defend agent-causation from popular objections—for example, objections advanced by Bernard Berofsky, C.D. Broad, Baruch Brody, John Bishop, Randolph Clarke, Roderick Chisholm, Donald Davidson, Alan Donagan, Carl Ginet, Ted Honderich, Robert Kane, John Thorp, and Gary Watson. I undermine the following two objections in detail. First, Peter van Inwagen argues that since an undetermined free action is impossible, no agent-causal theory of undetermined free action is correct. Second, Andrei Buckareff and Richard Feldman contend that agent-causalists cannot account for *rational* free action—that is, for an agent's acting freely *for* a reason.

Along the way, I defend Ted Warfield's novel argument for incompatibilism from Dana Nelkin and Samuel Rickless's recent criticism, I argue that Warfield's case fails for independent reasons. Nonetheless, I underscore the point that my agent-causal view best explains our pre-theoretic commitment that no directly free action could be determined.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to thank Al Plantinga and Tom Flint. Again and again, they gave my work a charitable reading and a thorough beating, a gift for which I'm deeply indebted. I likewise thank Peter van Inwagen, Fritz Warfield, and Mike Rea. As I reflect upon these five outstanding philosophers in whom I've found great inspiration, I hope to approach the rigor and clarity of their work and teaching styles.

A word of thanks extends to my friends at Sufficient Grounds Coffee House. They provided me with a fantastic space, hot Java, and stimulating conversations.

E.J. Coffman is the man.

Most of all, I thank my family. My wife, Jenn, has always provided me with constant support and encouragement. My hope is to be strong enough to return the favor. My eldest son, Eli, frequently gave me a boost of motivation both with his early-morning words, "Good luck on your diss-year-tation," and with his inquiries as to whether my dissertation was loud like unto a power tool. The tenderness of my daughter, Maeve, pushed me to get enough done each day so as to be home early enough to play. Finally, Silas Jonathan's kind spirit makes me all the richer.

CHAPTER 1

THE PROBLEM OF FREE WILL

This chapter introduces the topic of this book, which is free will. I outline a philosophical puzzle, the so-called metaphysical problem of free will. I follow Robert Kane, discussing a helpful taxonomy of conceptually distinct issues, each of which any satisfactory answer to the metaphysical problem of free will should consistently and informatively address. One may doubt whether anything new and interesting can be said about free will, given the list of impressive thinkers who have collectively grappled with various aspects of the problem of free will for over two millennia. I communicate a reasoned optimism about the prospects of further contribution to a corpus of literature arguably too large to master by a single philosopher. I then state my book's objectives and provide a skeletal overview of the remaining chapters. I conclude this chapter by noting several of my assumptions.

1.1   The Topic of Free Will
         Since free will is the object of our study, what intuitive notion do we have in mind with which we may begin? What do ordinary speakers of the English language have in mind when they speak about free will? At this point, we're not specifically interested in what philosophers think—we'll get to that soon enough. Instead, we aim first to grasp that notion of free will that is pre-philosophical, pre-theoretical, pre-analytical, and in a word, intuitive.[1]
         Virtually without exception as ordinary people entertain the question of whether or not they have free will, they without hesitation and without reservation believe that they do possess it. When asked to articulate what they mean, they usually reflect on their experiences involved in acting. It is not uncommon to hear ordinary folk claim that they *feel* free. Robert Nozick (1995: 101) addresses some of this phenomenology, noting:

> Making some choices feels like this. There are various reasons for and against doing each of the alternative actions or courses of action one is considering, and it seems and feels as if one could do any one of them. In considering the reasons, mulling them over, one arrives at a view of which reasons are more important, which ones have more weight. One decides which reasons to act on; or one may decide to act on none of them but to seek instead a new alternative since none previously considered was satisfactory.

I think most people would agree with Nozick's description of what frequently goes on, phenomenologically speaking, as one performs a paradigmatic deliberative action. Ordinarily, free will involves doing things that one wants to do. Acting freely typically involves doing

---

[1] Chapter 2 aims to get very clear about our pre-theoretic notion of free will.

things that one tries, chooses, or wills to do. One's free action, intuitively, cannot be forced to occur by something other than oneself. Instead, when acting freely, one is somehow a source—even an ultimate source—of something happening in the world. We believe that free actions are up to us. Freely selecting the cappuccino rather than the vanilla latté, for example, implies that the agent was both able in the present situation to select the cappuccino and able not to have selected the cappuccino (an ability that would be manifested by selecting the vanilla latté instead, perhaps). Picking the cappuccino over the vanilla latté was up to her in the sense that she at least had the power in the present situation to go either way.

According to the folk, then, many of our actions are very much like being able to order the cappuccino while also being able to refrain from ordering the cappuccino. Having power or ability to originate something that happens in the world is to have free will. Exercising this power to perform one of the alternative actions means that the action is free.[2] Nobody could in good faith doubt that many actions she performs are free. They are up to her.

1.2   The Metaphysical Problem of Free Will
        While the above discussion seems intuitive enough, certain theoretical considerations seem inconsistent with acting freely. Moreover, these theoretical considerations seem unreasonable to reject as well. The result is a puzzle that cries out for a solution. But what is the puzzle, broadly conceived?
        The problem is that free will seems incompatible with both determinism and indeterminism. Characterizing determinism with precision is no easy task.[3] In Chapter 3 we shall get very clear about our characterization of determinism. But for now, let van Inwagen's (1983: 2) and (1995a: 220) introductory construal suffice, "Determinism is quite simply the thesis that the past *determines* a unique future, that only one future is consistent with the past and the laws of nature."
        The problem, then, is that there are powerful arguments for thinking that free will is incompatible with determinism, and there are powerful arguments for thinking that free will is incompatible with *in*determinism. But since indeterminism is just the denial of determinism,

---

[2] Some philosophers would rather not use the term 'free will' at all. For example, Roderick Chisholm (1995: 98-99) says,

> I have not used the expression 'free will,' for the question of freedom, as John Locke said, is not the question 'whether the will be free'; it is the question 'whether the man be free' [John Locke (1894: Book II, Chapter xxi)]. The question is whether the agent is free to undertake any of those things he does *not* undertake and whether he is free not to undertake any of those things he *does* undertake.

Other philosophers are loath to use the term 'free will', use the term, but issue caveats. E.g., Peter van Inwagen (1990b: 288) says,

> The term 'free will,' when used in this sense, is a term of art. To ascribe 'free will' to an agent is simply to ascribe to the agent the property of having a free choice among certain alternatives. Such an ascription should not be taken to imply that the agent has a faculty called 'the will.'

And Jerome Weinstock (1976: 99) notes, "For when we talk of freedom of the will we are never talking about *the will* having power, but *the person* having certain (mental) abilities."

[3] Jordan Sobel (1998: 77-166) contends that there are at least ninety varieties of determinism.

they together exhaust the possibilities. So if free will is incompatible with both, then our conception of free will must itself be incoherent. However, surely free will is possible. After all, an agent's being morally responsible requires that she might possess and exercise her free will at least on some occasion. And surely it is possible for someone to be morally responsible. It follows that free will is possible.

A great deal, then, rests on the cogency of these powerful arguments mentioned above. Painting with broad strokes, what can we expect these arguments to look like?

Peter van Inwagen presents in a nutshell an argument for the incompatibility of free will and determinism. He dubs it the *Consequence Argument*, and it goes like this.

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.[4]

The basic idea is that it is impossible for a free action to be causally determined by preceding events none of which is up to the agent. We have no ability to affect causally any events in the very distant past. We have no ability to affect causally what are the laws of nature. But if events in the past together with the laws of nature guarantee precisely which events occur at any other time, then we have no *ability* or *power* to affect causally anything that happens now or in the future. It would seem that everything we do is, in a manner of speaking, dictated or guaranteed by forces beyond our reach. Thus, we could not be free in a system where every event is causally determined by temporally preceding events. So, free will looks incompatible with determinism. Or so the argument goes.

But it is also alleged that free will looks incompatible with the denial of determinism, i.e., indeterminism. Why do some philosophers think that free will is incompatible with *in*determinism? The essential features of the family of arguments for the incompatibility of free will and indeterminism can be summarized as follows. If a free action is not determined by events immediately preceding the action, then the occurrence of the action is simply a matter of brute chance. There is a chance that the action occurs, and there is a chance that the action does not occur. The agent's having certain beliefs and desires right before she acts does not causally guarantee which action she performs, since the action is not determined by any such events. But if one's own possession of one's beliefs and desires do not causally guarantee which action one performs, then what in the world does? Nothing, allegedly. The action's occurrence is random. But no event that randomly occurs is up to an agent. Therefore, since an undetermined action is a random event, the undetermined action is not up to its agent. But free actions are up to their agents. Thus, undetermined actions cannot be free. Therefore, undetermined *free* actions are impossible. So, free will looks incompatible with indeterminism too.

For cases of undetermined actions, then, there's nothing else to say that would explain how her performing the one action rather than a different action is up to her. Presumably we can only say, "She acted so-and-so immediately after having mental states such-and-such." And that is an explanation no deeper than stating, "Her acting so-and-so randomly followed her having mental states such-and-such." But since free actions must be more than random occurrences or

---

[4] Van Inwagen (1983: v, 16, 56, 222).

random followings, undetermined actions cannot be free.   There must be a deeper explanation for free actions than simply that they happened out-of-the-blue or simply that their agents just happened to do them. So, undetermined actions cannot be free. So, free will looks incompatible with indeterminism. Robert Kane (2002b: 415) succinctly reviews this chain of reasoning:

> An event that is undetermined might occur or not occur, given the entire past. So whether or not it actually occurs, given its past, would seem to be a matter of chance. But chance events are not under the control of anything, hence not under the control of the agent. How then could they be free and responsible actions? If a different choice might have occurred given exactly the same past, then exactly the same deliberation, the same thought processes, the same prior beliefs, desires and other motives—not a sliver of difference—that led to an agent's favoring one option (say, in choosing to vacation in Hawaii rather than Colorado), might by chance have issued in the opposite choice instead. If such a thing happened, it would seem a fluke or accident, like an uncontrolled quantum jump in the brain, not a rational, free, or responsible action.[5]

The intuitive appeal of arguments against undetermined free action rests in large part on our inability to see how indeterminism could *help* in understanding how we act freely. An action's being causally *un*determined by prior events does not at all look like a positive power exercised by the agent who acts. Pre-theoretically, a free action occurs in virtue of the agent's exercising a positive power to determine herself exactly which action she shall perform. But throwing indeterminism into the mix only seems to decrease any positive power the agent might already have. Indeterminism, then, seems to be more the slave of liability than a requirement of ability. Indeterminism provides nothing in addition to what is already available under determinism. Or so the argument goes.

Now we can see the force behind what's called *the metaphysical problem of free will*. Either the thesis of determinism is true or it is false. If determinism holds, there is no free will. But if indeterminism holds, there is no free will. So either way, there is no free will. That is to say, the notion of free will must be incoherent. Moral responsibility, however, obviously requires the presence of being free to act, and surely there is moral responsibility—or, at the very least, moral responsibility is possible. It follows that free will is both possible and impossible. Upon a bit of reflection, then, we find ourselves in a philosophical pickle.

1.3   Dissecting the Problem: Three Questions

This section is designed to orient the reader, dividing the metaphysical problem of free will into more manageable conceptual problems.[6] These separate yet interrelated problems can be distilled into a few fundamental questions. Understanding these questions better sets the stage for understanding my objectives and thesis, which I present in §1.5.

_____

[5] See also Kane (2002c).

[6] How I circumscribe the conceptual problems in this section is inspired largely by Kane (1996a: 12ff).

4

The first main distinction within the metaphysical problem of free will involves whether or not freedom of every kind worth wanting is compatible with determinism. Following Kane, addressing this issue trades on how to answer two questions, the Significance Question and the Compatibility Question.

First, what kinds of freedom or free will are worth wanting? This is the *Significance Question*. 'Freedom' can obviously be interpreted in a large number of ways. There are various sorts of freedom. There is freedom from constraint, freedom from coercion, freedom from addiction, freedom from self-deception, and freedom from punishment. There is freedom of speech and freedom of the press. There are legal freedoms, and there are political freedoms. Given such diversity, doubting that there is a single sense of freedom that cuts across such conceptual boundaries seems reasonable enough. However, we can eliminate many sorts of freedom from our discussion. For example, from §1.1 and later in Chapter 2, we notice that common folk have a fairly uniform and thematic notion of freedom that is more commonly associated with the philosophical term 'free will.' Common folk and many philosophers as well take it as basic that whatever kind of freedom or free will ends up being relevant to our study, it will be one that moral responsibility requires. That is, if there is moral responsibility, then there is some kind of freedom worth having. To be clear, this is not to say that freedom requires moral responsibility.

These brief remarks should not be taken to settle which sense of 'freedom' or 'free will' is operative in the rest of my book. Rather, they are intended to motivate getting as clear as we can about what sort of freedom is at issue in the metaphysical problem of free will. I probe the nuances of the Significance Question in Chapter 2, surveying how various philosophers characterize that kind of freedom worth wanting. Chapter 2, then, investigates various ways experts of free will initiate the uninitiated.

The second question, intimately connected to the Significance Question, asks whether or not this valuable kind of freedom is compatible with determinism. This is the *Compatibility Question*. The thesis that free will is compatible with determinism is called *compatibilism*, and its adherents are called *compatibilists*. Compatibilism is either necessarily true or necessarily false. Compatibilists believing the (we safely assume) contingent thesis of determinism are called *soft determinists*. The thesis that compatibilism is false is called *incompatibilism*, and its proponents are called *incompatibilists*.

The Compatibility Question by far receives the most attention in recent literature on free will. Much of the focus has been on the Consequence Argument, upon which §1.2 touched. Chapter 3 of my study investigates the Compatibility Question in more detail, exploring some reasons for incompatibilism.

I'm an incompatibilist. Moreover, I don't think my rationality compels me to have a cogent argument for incompatibilism. Nonetheless, I find the arguments against compatibilism persuasive. So, dialectically speaking, I want to hold the more modest conclusion that it is reasonable to hold that free will is incompatible with determinism—I do not pretend to have a demonstration against compatibilism.

Chapter 3 discusses a recent argument for incompatibilism, finding it flawed in the final analysis. I conclude that even if one's pre-theoretic commitment to incompatibilism is mistaken, it's consistent with compatibilism that an undetermined free action is possible. It should become clear, then, that there is plenty of motivation for pressing beyond the Compatibility Question. Experts of free will should take seriously the task of trying to render intelligible the notion of an undetermined free action.

The first main issue of the metaphysical problem of free will, then, addresses whether or not freedom of every kind worth wanting is compatible with determinism. The second main problem presupposes that significant free will, i.e., that kind of free will most worth having, is incompatible with determinism. The essence of the second main problem lies in figuring out how one can best account for free will's compatibility with indeterminism. More precisely, can one discover a principled answer to the following *Intelligibility Question(s)*. Can we make sense of a free will—a free will worth having—that requires (or is at least consistent with) indeterminism? Is such a freedom coherent or intelligible? If the Intelligibility Question cannot be answered, then either free will is incoherent or else its nature is essentially mysterious and hopelessly obscure.

Recall that all incompatibilists believe that free will is incompatible with determinism. Some incompatibilists believe that free will is incompatible with *in*determinism as well.[7] Thus, they are committed to believing that free will is incompatible with itself, in which case free will is self-contradictory and therefore impossible. Other incompatibilists are agnostic about the possibility of free will. The remaining incompatibilists, and by far the majority of them, believe that free will is possible.[8] Some incompatibilists hold that determinism is in fact true and therefore that free will does not in fact exist.[9] They are called *hard determinists*; their position is called *hard determinism*. Most incompatibilists, though, believe that most of us average folk actually have free will. These incompatibilists are called *libertarians*. Their position, *viz.*, that free will is incompatible with determinism and someone at some time acts freely, is called *libertarianism*.

Thus, the Intelligibility Question is pressing for philosophers who think libertarianism is possible. For, if one takes the metaphysical problem of free will seriously, arguing for the incompatibility of free will and determinism is not enough. Experts of free will who believe that free will non-vacuously implies indeterminism have their work cut out for them. Their aim should be to provide a reasoned response to the Intelligibility Question, rendering comprehensible the claim that free undetermined actions are possible. Moreover, since most incompatibilists are libertarians, and since most libertarians believe that nearly all of us have freedom, they should be able to account for how our own free deliberative (and rational) actions are compatible with—and indeed require—their being undetermined.

As we've seen, one can divide the metaphysical problem of free will into simpler, slightly more manageable problems. Their solutions would just be answers to the following questions.

- Is there any kind of freedom worth wanting that is incompatible with determinism, and if so, what kind is it? (Significance Question)

- Why is it incompatible with determinism? (Compatibility Question)

- Can we make sense of a free will that is incompatible with determinism? Is such a freedom coherent or intelligible? (Intelligibility Question)

---

[7] E.g., see Galen Strawson (1986: 28-29) and Richard Double (1991).

[8] For some specimens in the minority camp, see van Inwagen (2001, 2000), Alicia Finch (2002), and Thomas Nagel (1986).

[9] E.g., Paul Holbach (1850) and Ted Honderich (1993).

1.4   Nothing New Under the Sun?
       Susan Wolf (1990: vii) writes,

> Free will is arguably the most difficult problem in philosophy. Given the
> centuries of thought, even of deep and brilliant thought, that have been devoted to
> this problem, it would not be unreasonable to assume that only fools rush in, at
> this point, thinking they have something to say about it.

Wolf recognizes that the written precedent set by philosophers, theologians, and other scholars
thinking on the topic of free will is enough to overwhelm anyone interested in understanding all
of their insights.[10] Moritz Schlick (1939: 143) opines:

> …[I]t is really one of the greatest scandals of philosophy that again and again so
> much paper and printer's ink is devoted to [the so-called problem of the freedom
> of the will], to say nothing of the expenditure of thought, which could have been
> applied to more important problems (assuming that it would have sufficed for
> these).

So, why spill more ink? Why think that anything new and interesting could still be said? Given
the sheer volume of the work already published and given the impressive credentials of its
contributors, any prospective contributor should consider very carefully the responsibility to
avoid further bogging down the literature on free will.

This problem is not unique to discussions of free will but to philosophy in general. The
same sort of worry could just as easily be made with respect to other domains of philosophical
inquiry, such as meta-ethics or epistemology or the problem of universals. However, the sting of
the charge remains. Are there any specific reasons for thinking that there are fertile conceptual
pockets for today's philosopher, a philosopher seeking to clarify, strengthen, solve, or dissolve
the metaphysical problem of free will? I think so.

Until quite recently most scholars, perhaps under the residual influence of Thomas
Hobbes, believed one's free action could be determined. Three to five decades ago there were
many philosophical pieces that contested this stance on the problem of free will and

---

[10] Arguably, Aristotle held that freely performed actions cannot be the inevitable consequences of
factors lying in our deep past, long before we even attempt to perform any actions—e.g., see Richard
Sorabji (1980: 233ff). The problem between concepts of determinism and free will, however, was first
given systematic treatment by the Epicureans and then by the Stoics. The Epicureans contended that
one's performing an act within a deterministic system is impossible. They postulated that somewhere in
the otherwise deterministic activity of the atoms constituting us, an atom randomly swerves and thereby
grounds our freedom. The Stoics opposed this view, contending that one can act freely even if
environmental factors antecedently guarantee one's every move. For a most excellent translation and
commentary on the Epicurean and Stoic views regarding determinism and free will, see David Sedley and
Anthony Long's (1999: 102-112, 386-394). Other remarkable philosophers grappling with issues in free
will include Saint Augustine (1964), Saint Thomas Aquinas (1945), William of Ockham (1983), Gottfried
Wilhelm Leibniz (1988, 1951), Thomas Hobbes (1962), Samuel Clarke (1738), Locke (1894), David
Hume (1739), Thomas Reid (1895), and John Stuart Mill (1947). More recently, see Alfred Ayer (1954)
and Bertrand Russell (1953). For an impressive bibliography, see Kane (2002a: 577-617).

determinism.[11] Nonetheless, we do not encounter anything like a rigorous version of the Consequence Argument (mentioned above) until the advent of articles such as David Wiggins's (1973) 'Towards a Reasonable Libertarianism' and van Inwagen 's (1975) article 'The Incompatibility of Free Will and Determinism.' Van Inwagen further developed his view, which culminated in his (1983) book, *An Essay on Free Will*. There was a surge of interest in replying to the incompatibilist project van Inwagen clearly underscored, and then of course there were the replies to the replies. As the epicycles continue to this day, many philosophers propose that the debate has reached an impasse, adherents on each side digging in their heals.[12] However, these exchanges are less than two decades old! It's reasonable to think that philosophers have not exhausted everything new and interesting on this front.[13]

Moreover, there is an underrepresented but growing interest attempting to articulate rigorously how an undetermined action could nonetheless be up to the agent who performs it.[14] These perspectives, many of which are no more than a decade old, reject the arguments concluding that free will is incompatible with indeterminism. If there is room for novelty, I believe there is space here. I share the same conviction expressed by Kane (2002b: 406),

> It was often said in the twentieth century that the free will issue is a "dead issue." All the passages in the labyrinth had been traveled and retraveled. Since I first began thinking about this topic thirty years ago, my conviction, to the contrary, was that whole passages in the labyrinth of free will remained unexplored or others were too lightly explored.

1.5   This Book's Objectives Stated

My primary interest lies in evaluating several of the best responses to the Intelligibility Question. I would like to move directly to discussing those views seeking to render intelligible how there could be free will that requires indeterminism—or, at the very least, how there could be an undetermined free action. However, as noted in §1.3, answering the Intelligibility Question presupposes a more careful inspection of our intuitive grasp of what free will is and why it is so important. And surveying positions with respect to the Compatibility Question seems appropriate as well, as some philosophers become interested in the Intelligibility Question

---

[11] E.g., M.R. Ayer (1968), C. D. Broad (1952), Roderick Chisholm (1958, 1976), Austin Farrer (1967), Carl Ginet (1966), Anthony Kenny (1975), J.R. Lucas (1970), A.I. Melden (1961), and Richard Taylor (1992).

[12] E.g., see Kane (1996a: 14, 16) and Timothy O'Connor (1995a: 4) and (2000a: 5, 17).

[13] E.g., see Thomas Crisp & Ted Warfield (2000), Alicia Finch & Ted Warfield (1998), Warfield (2000), Dana Nelkin (2001), Dana Nelkin & Samuel Rickless (2002), and O'Connor (2002b).

[14] Most notably, see Randolph Clarke (2003, 2002, 2000, 1999, 1997, 1996a, 1996b, 1995, 1993), William Craig (1991), Laura Ekstrom (2001b), Thomas Flint (1998a), Carl Ginet (2002, 1997, 1996, 1995, 1990, 1983), Stewart Goetz (2000, 1997, 1988), William Hasker (1999, 1995, 1989), Kane (2002a, 2001, 2000a, 2000b, 1999a, 1999b, 1999c, 1996a, 1996b, 1988, 1985), O'Connor (2002b, 2000a, 1996, 1995a, 1994, 1993a), William Rowe (1991a, 1989, 1987), Edward Wierenga (1991, 1989). Virtually none of these pieces is more than a decade old, and it would be quite unreasonable to think everything new and interesting has been said regarding these fresh contributions.

because, first, they find convincing certain arguments for incompatibilism and *then*, second, they see that indeterminism appears to threaten free will just as much as determinism does. So, investigating arguments for incompatibilism benefits the present study, as they play a large part in motivating one's tackling the Intelligibility Question in the first place.[15] A survey of most work in free will over the past few decades suggests that serious advances with regard to the Compatibility Question are becoming scarcer. Nonetheless, I think philosophers are still crafting rigorous and credible answers to the Intelligibility Question.

Some philosophers who accept incompatibilism, however, believe that answering the Intelligibility Question may be preempted by certain direct arguments for the conclusion that free will is either self-contradictory or impenetrably mysterious. Van Inwagen, for example, believes this. Chapter 4 finds van Inwagen's recent line of thought inadequate.[16] If I'm right, there is one less barrier to taking seriously prospective answers to the Intelligibility Question.

There are three major approaches to developing positive replies to the Intelligibility Question. In keeping with recent terminology, I call these views *simple indeterminism*, *causal indeterminism*, and *agent-causation*. Chapter 5 surveys them in detail. My discussion of simple indeterminism and causal indeterminism involves criticizing a strong representative (or two) of each approach. I criticize two of the best theories of simple indeterminism (Ginet and Goetz's), and I criticize one of the best theories of causal indeterminism (Kane's). Chapter 5 concludes by clearly describing the essential features of Timothy O'Connor's (2000a) theory of agent-causation, which is the view I find most promising.

The remaining four chapters seek to defend and to some degree extend O'Connor's view. Chapter 6 discusses the concepts of causal production and its genus, partial causation or causal contribution. I modify and defend a version of Roderick Chisholm's (1986) account of causal contribution. I contend that wedding my version of causal contribution to O'Connor's theory results in a sophisticated agent-causal theory that parries some of the most pressing criticisms against O'Connor's view. Chapter 7 works toward earning the truth of this contention by defending O'Connor's theory of agent-causation from many popular objections. Chapter 8 addresses perhaps the most damaging objection against agent-causal views, *viz.*, that they cannot account for how an agent's possessing reasons enters into part of an explanation for why the agent freely acts as she does. In particular, I undermine the most recent objection against O'Connor's sufficient conditions for reasons-explanations.[17] I conclude that agent-causation is the most plausible theory of free will.

---

[15] I should emphasize that one could be a compatibilist and still seek to answer a similar version of the Intelligibility Question, *viz.*, can we make sense of a free will that is compatible with *in*determinism—is such a freedom coherent or intelligible? An affirmative answer to this version of the question is implied by but does not imply an affirmative answer to the original Intelligibility Question. That is, if a free will worth having is both coherent and requires indeterminism, then free will is consistent with indeterminism. Some compatibilists may hold that, in addition to the possibility of a determined free action, there could be undetermined free actions as well—see Alfred Mele (1995).

[16] Van Inwagen (2001, 2000, 1997, 1993).

[17] Richard Feldman & Andrei Buckareff (2003). My treatment applies equally well against Randolph Clarke's (2003) version of Feldman & Buckareff's objection.

In fine, I aim to defend an agent-causal approach to the Intelligibility Question.[18] I defend O'Connor's theory—or a close cousin of it—by (i) refuting a popular argument for the impossibility of free will, (iv) showing how agent-causation is superior to other theories of undetermined free action, (iii) buttressing an agent-causal theory with a rigorous account of causal contribution, and (iv) disabling some of the most popular objections against agent-causation.

1.6   Methodology and Other Assumptions
        While not much is taken for granted among analytic philosophers, the present study's scope requires making many presuppositions. Adopting reasonable assumptions from the outset is nothing new. Consider The Philosopher's counsel:

> We must, as in all other cases, set the phenomena before us and, after first discussing the difficulties, go on to prove, if possible, the truth of all the reputable opinions about [our topic] or, failing this, of the greater number and the most authoritative; for if we both resolve the difficulties and leave the reputable opinions undisturbed, we shall have proved the case sufficiently.[19]

For Aristotle, the reputable opinions consist of what is "…accepted by everyone or by the majority or by the wise—i.e. by all, or by the majority, or by the most notable and reputable of them."[20] It is worth emphasizing that Aristotle holds that what everyone believes (the *endoxa*) cannot rationally be challenged. He asserts, "…[W]e say that that which everyone thinks really is so; and the man who attacks this belief will hardly have anything more credible to maintain instead."[21] Thomas Reid (1895: 29) reiterates the point, saying,

> For though, in matters of deep speculation, the multitude must be guided by the philosophers, yet, in things that are within the reach of every man's understanding, and upon which the whole conduct of human life turns, the philosopher must follow the multitude, or make himself perfectly ridiculous.

Aristotle and Reid realize that doing philosophy begins by gathering what has been accepted, raise puzzles for it, solve the puzzles as far as possible, and retain all that survives the accepted opinions.[22] Alvin Plantinga (1990: 33) seems to agree with Aristotle regarding philosophy's point and point of departure, saying:

---

[18] O'Connor (2002a, 2002b, 2000, 1996, 1995b, 1993a).

[19] *Nichomachean Ethics* (*NE*) VII 1, 1145b1-7. Compare *Topics* I 1, 100a21-23.

[20] *Topics* I 1, 100b21-2.

[21] *NE* X 2, 1173a2-3; cf. *Physics* 196b14.

[22] *NE* VII 1, 1145b2-7 and 1146b6-8. See Sorabji (1980: 246). Consistent with Aristotle's dialectic method of doing philosophy, Frank Jackson (1998) sketches a general method for how a philosopher does (or ought to do) serious metaphysics. He argues rather convincingly that the process of doing metaphysics includes taking any putative feature of the world and then either eliminating the

Philosophy is in large part a clarification, systematization, articulation, relating and deepening of pre-philosophical opinion. We come to philosophy with a range of opinions about the world and humankind and the place of the latter in the former; and in philosophy we think about these matters, systematically articulate our views, put together and relate our views on diverse topics, and deepen our views by finding unexpected interconnections and by discovering and answering unanticipated questions. Of course we may come to change our minds by virtue of philosophical endeavor; we may discover incompatibilities or other infelicities. But we come to philosophy with pre-philosophical opinions; we can do no other.

So we enter into philosophical inquiry with pre-analytic beliefs and convictions in hand. The next natural question concerns *who* gets to count as (or as among) the most authoritative, most notable and reputable, and wise. I do not have a principled resolution to this question. However, I'm quite skeptical that counting noses settles the matter. Plantinga is too. For example, though most philosophers are not Christians, Plantinga contends that the Christian philosopher has every right to take as primitive what the Christian philosophical community believes, unless there are cogent arguments to the contrary, where these arguments have premises that a Christian should believe. Plantinga (1990: 33) concludes:

> The Christian has as much right to her pre-philosophical opinions as others have to theirs….[If the Christian philosopher's pre-philosophical opinions] are widely rejected as naïve, or pre-scientific, or primitive, or unworthy of "man come of age," that is nothing whatever against them. Of course if there were genuine and substantial arguments against them from premises that have some legitimate claim on the Christian philosopher, then she would have a problem; she would have to make some kind of change somewhere. But in the absence of such arguments— and the absence of such arguments is evident—the Christian philosophical community quite properly starts, in philosophy, from what it believes.

Plantinga goes so far as to argue that many uniquely Christian commitments are not only opinions but count as knowledge as well.[23]

The upshot is this. I endeavor to take seriously both the common folk's pre-analytic beliefs and convictions as well as the possibility of Christian beliefs. Moreover, when deploying additional assumptions later in my study (e.g., in Chapter 2), I do my best to respect the relevant pre-analytic beliefs of the wise, most notable, and most reputable among us. I aim, then, to outline the common beliefs with which many philosophers begin their studies of free will. I consider assumptions thus gathered to be innocent until proven guilty. That is, I assume they are

---

feature altogether—which means that the given feature was merely putative—or locating the given feature, where the latter is usually done by identifying it with some other, more fundamental, feature of the world. If the original putative feature can neither be eliminated nor located via identification with a more fundamental feature, then the putative feature just is a fundamental feature. The result is a metaphysical theory that concerns both all and only things that exist as well as their natures, i.e., as well as what the things are like.

[23] Plantinga (2000, 1993a, 1993b).

true until shown false with an obviously valid argument whose premises are less controversial than the denial of its conclusion. My study's primary objective, then, is to defuse the problem of free will while keeping as many of our examined assumptions intact.[24]

In the balance of this chapter, I assemble some peculiarly philosophical assumptions. Every philosopher discussed in this book adheres to them—or at least deems them reasonable enough to tolerate them. They are as follows.[25]

I assume that there are *propositions* (e.g., that Colorado has no natural lakes, or that computers deserve praise). Propositions are non-linguistic bearers of truth-values. They must be either true or false, but propositions cannot be both true and false. I assume that every false proposition nonetheless exists. Say that one proposition *p entails* another proposition *q* just in the case that it is impossible for *p* to be true while *q* is false.

I assume there are states of affairs, where a state of affairs is simply a way things might be or might not be (e.g., the beetle's being black, a kangaroo's having no tail, or a triangle's having four sides). A state of affairs must either obtain or fail to obtain, but a state of affairs cannot both obtain while failing to obtain. A non-obtaining state of affairs nevertheless exists. Say that a state of affairs *S includes* a state of affairs *S** precisely when it is impossible for *S* to obtain while *S** fails to obtain. Say that a state of affairs *S precludes S** if and only if it is impossible that both *S* obtains and *S** obtains. Say that two states of affairs are *consistent* if and only if one does not preclude the other.

I assume that propositions and states of affairs have modal features. For example, I assume that a proposition is *necessary* (i.e., necessarily true) just when it is impossible for it to be false. A state of affairs is necessary just when it is impossible for it to fail to obtain. A proposition is *impossible* if and only if it is necessarily false. A state of affairs is impossible if and only if it necessarily does not obtain. A proposition or a state of affairs is *contingent* just when it is neither necessary nor impossible. A proposition is possible if and only if it is not necessarily false, i.e., it might be true. A state of affairs is possible if and only if it does not necessarily fail to obtain, i.e., it might obtain.

I assume that there are possible worlds. A possible world is a complete way things might be. That is, a possible world is a state of affairs *S* that possibly obtains and is such that for any state of affairs *S**, *S* either includes *S** or precludes *S**. Possible worlds, then, are maximally consistent states of affairs. The actual world is just an obtaining state of affairs that includes every obtaining state of affairs. One may also think of a possible world as a self-consistent proposition *p* either entailing *q* or entailing ~*q* for any proposition *q*. On this interpretation of a possible world, the actual world is just the proposition that entails all and only those propositions that are actually true.

Possible worlds are useful in characterizing the modal status of propositions or states of affairs. For example, let a proposition be true in a possible world just in the case that the proposition would be true if that possible world were actual. A possibly true proposition is a proposition that is true in a possible world. For any proposition *p*, let '*p*-world' designate a possible world in which p is true. A necessary proposition is one that is true in every possible world, and an impossible proposition is true in no possible world. Given the apparatus of

---

[24] For more on our common belief's being innocent until shown mistaken, see Richard Taylor (1992: 2-3).

[25] For more on these assumptions, see Plantinga (1974: Chapters 1-6) and Wierenga (1989: 6-11).

possible worlds, we can give another characterization of the logical relation of entailment. A proposition p entails a proposition q if and only if every p-world is a q-world.

I assume that there are properties (e.g., being six units in length, or being round), and that every property exists even if there is nothing that has it (e.g., being taller than the tallest object). I assume that there are individual substances, where an individual substance is an object that has a property but this object is neither a proposition, a state of affairs, a property, nor any combination of these. I assume that a substance exists in a possible world W just if it is the case that if W were actual, then the substance would exist. I assume that every substance has some of its properties essentially, where a substance Δ essentially has a property P just in the case that, necessarily, if the Δ exists, then Δ has P. For example, suppose that there is a substance named Sam. The property *being someone who exists* is a property essentially had by Sam. For, every world in which Sam exists, Sam bears the property *being someone who exists*.

Finally, I assume that there are true contingent counterfactuals. A counterfactual is a proposition of the form "If it were the case that p, then it would be the case that q," for any propositions p and q. Normally, counterfactuals are expressed in the subjunctive mood. Let "$\Box\to$" denote the counterfactual connective. Accordingly, let "$p \Box\to q$" abbreviate "If it were the case that p, then it would be the case that q." Say that a counterfactual ($p \Box\to q$) is true at a world W if and only if either (i) there is no p-world, or (ii) some ($p\&q$)-world is more similar than any ($p\&\sim q$)-world to W.[26] For example, consider the counterfactual *If all beer were non-alcoholic, then fewer people would drink beer*. The antecedent of this counterfactual happens to be contrary to fact, i.e., it is false. But the proposition *all beer is non-alcoholic* is true in some possible world. Moreover, some possible world where all beer is non-alcoholic *and* fewer people drink beer is closer to our world than any possible world where all beer is non-alcoholic and just as many people[27] drink beer. Intuitively, the degree of departure from reality to get a world in which all beer is non-alcoholic and fewer people drink beer is less than the degree of departure from reality to get any world in which all beer is non-alcoholic and just as many people drink beer. The three-place comparative similarity (or closer-than) relation between possible worlds is taken as primitive.

Now that the reader is familiar with the metaphysical problem of free will, my objectives, and several of my assumptions, let's turn to getting clearer about how philosophers construe the subject of our study. Results from the next chapter continue to grow our set of assumptions by consulting what the wise among us have to say about free will.

---

[26] See David Lewis (1986) and (1973).

[27] That is, just as many people as the number of beer drinking people in our world, the actual world.

CHAPTER 2

THE SIGNIFICANCE QUESTION

*Free will is the power of agents to be the ultimate creators (or originators)
and sustainers of their own ends or purposes.*[1]
—Robert Kane

Taking our cue from our discussion of Aristotle in §1.6, we further fix our study's foundation by outlining that which is accepted by everyone or by the majority or by the wise—that is, by all or by the majority or by the most notable and reputable of them. In §1.1 we described in a dearth of detail what common people believe about free will. Here, I extend that description in two ways.

First, I delve into what many of the experts have to say about free will. Getting clearer about what we mean by 'free will' better ensures that the rest of our study is not frivolous. For, if we fail to solidify more carefully what we mean, then we cannot satisfactorily answer questions such as *Is free will compatible with determinism?* or *Does argument so-and-so provide good reason to think that free will is impossible?* or *Does having feature such-and-such require that one act freely?* or *How can someone act freely for a specific reason that she has?* Paul Benson (1987: 465) explains:

> Without some initial, relatively fixed beliefs about free action to start with, we would not know what we are theorizing about. And, apart from questions concerning the subject of our theory, we would not have any motivation to theorize; nor would we have much of a clue about how to begin. Part of the reason why free action has been the occasion for such controversy and confusion is that philosophers have not worked hard enough to display the intuitive origins and motivating questions for their favored theories of free action.

We'll discover that many philosophers share a strict and demanding, yet intuitive, notion of free will. More precisely, it will become clear that acting freely requires two conditions: (i) the agent has both a power to act a certain way as well as a power not to act in that way, and (ii) that the agent herself is an ultimate source of change in the world. The next chapter, Chapter 3, sketches reasons for thinking that this demanding kind of freedom is incompatible with determinism, which sets the stage for the difficult task (addressed in still later chapters) of trying to defend the claim that this demanding kind of freedom is possible.

Second, I entertain reasons for why this kind of freedom matters to us in the first place. I conclude that there is a strong case for thinking that many valuable features of our lives either require or are required by this strict and demanding kind of freedom.

---

[1] Kane (1996a: 4, 15, 78, 196).

This chapter, then, carefully unpacks how various experts intuitively characterize our pre-theoretic notion of free will. It clarifies our intuitions about what acting freely is and why it is desirable. After understanding that there is a fairly well-behaved and deeply significant concept of free will both shared by many experts and consistent with folk beliefs, we will be in a better position to evaluate the arguments of Chapter 3. Chapter 3, as you may recall, investigates whether or not this significant kind of freedom is compatible with determinism.

2.1   The Easy Road and the Hard Road

Roderick Chisholm issues a healthy warning, claiming to underscore what our pre-theoretical notion of freedom is <u>not</u>. He highlights a distinction that Thomas Aquinas draws between the *actus voluntatis imperatus* and the *actus voluntatis elicitus*. The former concerns whether a person is free to bring about what it is that she undertakes, or wills, to bring about. The latter concerns whether a person is free to undertake, or to will, the things that she undertakes, or wills.[2]

To illustrate this distinction, Chisholm considers the question 'Is the person free to bring about what it is that she wills (or undertakes) to bring about?' He (1995: 99) writes, "*This* question is not difficult to deal with. We may answer it affirmatively by pointing out that on occasion people do do the things that they will [or undertake] to bring about." Suggesting that the problem of free will cannot be so easily dismissed, Chisholm concludes that those neglecting the above distinction by focusing only on the easy question "bypass the more fundamental question of freedom." The harder, more fundamental question of freedom is this: 'Is a person free to undertake/will what it is that she undertakes/wills?'

I suggest that we take Chisholm's advice seriously. I see no good reasons not to take him seriously. Of course, this does not imply that there is no good reason. (It would be nice were our good-reason-detection-faculties sufficiently calibrated.) However, the question 'Is a person free to undertake what it is that she undertakes?' seems straightforwardly intelligible. And if it is intelligible and might be answered affirmatively, then it seems to follow that any comprehensive solution to the problem of free will should answer (or at least leave room for) how this could be. However, if we take Chisholm's advice seriously, then many accounts of freedom are too narrow in that they fail even to address how an agent can freely undertake what it is that she undertakes.

For example, Thomas Hobbes states, "Now he that reflects so on himself cannot but be satisfied…that a *free agent* is he that can do as he will, and forbear as he will, and that liberty is the absence of external impediments."[3] Or consider David Hume (1955) who says,

> By liberty [i.e., free will], then, we can only mean *a power of acting or not acting, according to the determinations of the will*; that is, if we choose to remain at rest, we may; if we choose to move, we also may.

---

[2] Chisholm (1995: 99), referring to Aquinas' *Summa Theologica*, First Part of the Second Part, Question I, Article 1. See also Chisholm (1976a: 66).

[3] Quoted in Honderich (1993: 96).

John Locke falls into the same broad camp, characterizing freedom as "the power a man has to do or forbear doing any particular action according…as he himself wills it."[4] According to Rowe's (1991a: 2) more polished interpretation of Locke's construal, one performs action A freely just in the case that "one has both the power to do A should one will to do A and the power to refrain from doing A should one will to refrain." G.E. Moore (1911: 211) apparently agrees, stating:

> There is, therefore, much reason to think that when we say that we *could* have done a thing which we did not do, we *often* mean merely that we *should* have done it, *if* we had chosen. And if so, then it is quite certain that, in *this* sense, we often really *could* have done what we did not do… And for my part I must confess that I cannot feel certain that this may not be *all* that we usually mean and understand by the assertion that we have Free Will.

The apparent difficulty with each of these views is that none accommodates the intuition that our undertakings, willings, or choosings themselves can be free. That is, each of these views fails to capture how one could be free in her very act of willing or in her very act of undertaking. Locke quips, "For how can we think any one freer than to have the power to do what we will."[5] Consonant with Chisholm's advice, Bishop John Bramhall answers plainly, "…[T]he will of man is free, not only to do if he will, but to will, that is, to choose or to elect."[6] Even Harry Frankfurt—who has arguably done as much as any contemporary philosopher to deepen compatibilist analyses of free will—agrees, asserting, "When we ask whether a person's will is free we are not asking…whether he is free to do as he pleases."[7]

We thus take seriously the possibility that an agent is free to undertake what it is that she undertakes—or, equivalently (I assume), that she is free to will what it is that she wills, that she is free to choose what she chooses. This is not to throw out the baby with the bath water. We may still concede to Hobbes, Hume, Locke, and Moore that certain interruptions between our undertakings and our undertakings being carried out, say, to overt actional fruition reduce the number of free actions. However, we also reasonably propose that there is more to freedom than a lack of impediments between one's undertaking and the act for which the undertaking is made.

2.2   What Many Philosophers Say about Our Pre-theoretic Notion of Free Will

So how do philosophers initiate the uninitiated into the complexities of free will? How do the wise characterize our pre-philosophical conception of free will? Gary Watson, who is no friend of libertarianism, contends that "any reasonable conception" of free will must capture the

---

[4] Locke (1894: Book 2, Chapter 21, §15).

[5] Locke (1894: Book 2, Chapter 21, §21).

[6] Bramhall (1844: volume 4, 258).

[7] Frankfurt (1988: 20). Though the accused philosophers above are compatibilists, one should not take my remarks in this paragraph to count as an argument against every compatibilist account of free will. It is an open question, at least at this point, whether there is an adequate compatibilist account.

right interpretation "of two different features of freedom…—namely, self-determination (or autonomy) and the availability of alternative possibilities."[8]

I find Watson's claim insightful. According to nearly every expert of free will, acting freely implies that the agent has the power to do otherwise than she actually does. Even more precisely, acting freely implies that the agent has the power not to do what she actually does. The agent, shall we say, has dual ability. Moreover, and no less importantly, acting freely implies that the agent *herself* ensures which of the available courses of action materializes. The agent is self-determining, somehow being an ultimate source or originator of some change in the world. The rest of this section provides evidence for thinking that many philosophers assume these two features of freedom.

2.2.1   Dual Ability
John Searle (1984: 95) claims,

> In normal behaviour, each thing we do carries the conviction, valid or invalid, that we could be doing something else right here and now, that is, all other conditions remaining the same. This, I submit, is the source of our own unshakable conviction of our own free will.

Laura Ekstrom (2000: 3) describes one's free action as being "up to oneself." Intuitively, a free act is within the agent's power. When acting freely, the agent exercises her power to so act. Regarding these features of being up to us and being within our power, Aristotle claims that "…when acting is up to us, so is not acting…" and "…where it is in our power to act it is also in our power not to act, and *vice versa*."[9] Aristotle sees that when our actions are up to us we have (what I shall call) dual power or dual ability.

An agent has *dual ability* or *dual power* precisely when she acts a certain way yet had it within her power not to act in that way. Having dual power, then, means that the agent is able to perform an action *and* is able to refrain from performing that action. Dual power, then, is modal. When exercised, it attaches not only to what the agent does but also to what the agent *could* do instead. This conception of dual power or dual ability is not to be confused with being able to perform two (or more) actions either at once or in sequence. Acting with dual power indicates that up until the agent acts, she has it within her power to refrain from so acting.

Richard Double, who incidentally believes that there is no coherent notion of free will, claims that compatibilists and incompatibilists alike have reason to believe that free will requires having dual ability. Double (1991: 12) claims that free will implies "that free agents have the ability to choose and to act differently than they actually do." "The point here," Double (*ibid.*) continues, "is that free agents do not have to make the choices they do; they have the ability…to choose otherwise." The point here is that even Double recognizes a conception of freedom that implies having dual ability.

---

[8] Watson (1987: 145). "Even classical compatibilism," Watson (*ibid.*) asserts, "made room, or tried to make room, for both." Etymologically, the term 'autonomy' means self-rule or self-government.

[9] Aristotle, *NE* III 5, 1113b6-8.

17

Thomas Reid (1895: Essay 4, Chapter 1) claims that a person is free at some time only when she has power over the determinations of her own will at that time. Her action is free at least partly in virtue of her exercising this power in producing her action. According to Reid (1895: Essay 1, 65), a person's having the power to perform an action A implies that she has the power not to perform A. Consequently, Reid believes that free will requires dual ability.

Carl Ginet concurs. He (1995: 69) claims, "By a *free* action I mean one such that until the time of its occurrence the agent had it in her power to perform some alternative action (or to be inactive) instead." Randolph Clarke (2003: 15) reiterates the same basic idea, saying,

> When an agent acts with free will, she is able to do other than perform the action that she actually performs; she has a choice about what whether she performs the action; what she does is up to her.

Van Inwagen (1995a: 221) notes that philosophers may describe acting with free will in language that is not ordinary English at all, saying,

> They may, for example, talk of a sheaf of alternative [i.e., incompatible] possible futures that confront the agent, and say that he has free will just in the case that more than one of these futures is 'open' to him or 'accessible' to him.

Does this remark imply that free will requires having dual power? Yes. For, an agent accesses a future previously open to her precisely when she acts a certain way, which manifests her ability to act in that way. Moreover, at least one other future is open to her (at least partly) in virtue of her ability not to act as she actually does. Thus, she has dual power. Concerning this property *being open*, Ginet (1990: 9) says,

> Two or more alternatives are *open to me* at a given moment if which of them I do next is entirely up to my choice at that moment: Nothing that exists up to that moment stands in the way of my doing next any one of the alternatives.

Honderich (1988: 385) states that "…our actions are such that our futures are in part *open, alterable, or unfixed*." A future's being open to an agent (or alterable, or unfixed), according to this more technical philosophical nomenclature used by van Inwagen, Ginet, and Honderich, implies that the agent has dual power.

Van Inwagen, of course, does not introduce the uninitiated into the complexities of free will with technical philosophical jargon. The jargon comes later. Rather, he first focuses on the locution 'free will', suggesting that it normally occurs only in philosophical or theological discussions. He notices that, in the current analytical tradition, having free will does not presuppose any special faculty called the will whose operations are free when unimpeded. Rather, according to van Inwagen, one's having free will *just means* that one is in a position where one can or could do otherwise. One might inquire about what it means for Mrs. Thatcher, for example, to have free will. Van Inwagen (1995a: 220) asserts that it means that she is at least sometimes in a situation like the following: "She is contemplating incompatible courses of action A and B (lecturing the Queen and holding her tongue, say), and she *can* pursue the course

of action A and *can also* pursue the course of action B."[10] In other words—and in propositions perhaps neither more nor less informative when considered individually—she is able to do A and is able to do B, or she has it within her power to do A and has it within her power to do B. The notion of dual power, then, lies at the heart of our conception of free will.

Van Inwagen contends that the keystone of debates in free will lies in discovering the correct philosophical interpretation of the phrase 'could do otherwise.' Blunders arise, van Inwagen warns, in thinking that there are different meanings of 'could do otherwise.' For example, it is a misunderstanding to believe that there is a compatibilist sense of 'could do otherwise' and an incompatibilist sense of 'could do otherwise.' All of the disputing parties *mean* the same thing by this phrase, which is the meaning afforded by ordinary language.[11] Otherwise, van Inwagen reasons, there would be no substantive disagreement between compatibilists and incompatibilists. According to van Inwagen, their disagreement concerns the single correct interpretation of 'could do otherwise.' In other words, they disagree about the precise conditions for acting with dual power. Double (1991: 12) appears to agree, saying:

> That compatibilists and incompatibilists are committed to the view that any acceptable sense of 'free will' must entail that free persons could have chosen otherwise is shown by the extensive twentieth-century debate over whose account of this notion captures 'the' prephilosophical notion (Moore, 1911; Austin, 1961; Chisholm, 1976; Aune, 1967; Lehrer, 1968).

I find van Inwagen's caveat insightful. Nonetheless, perhaps the scope of its application should not be extended too far. The phrase 'could do otherwise' occurs in ordinary language, and van Inwagen (1995a: 220) recognizes that 'free will' normally does not occur in normal discourse—modulo the clause 'acting of one's own free will.' Nonetheless, van Inwagen's contention that there is but one sense of 'could do otherwise' is consistent with there being multiple senses of 'free will', 'has free will', 'acts with free will', 'acts freely', and the like.

Ted Honderich (1993: 103) asserts, "We don't have *any* definition of a free choice if a definition is supposed to be the one and only correct description of a thing."[12] Honderich believes that 'free' and kindred terms are systematically ambiguous. On the one hand, they designate voluntary choices and actions. On the other hand, they designate voluntary *and originated* choices and actions. In contrast, van Inwagen (1995a: 220, 1983: 8) merely stipulates but does not argue for the claim that an agent's acting with free will at a particular time *just means* that the agent could do otherwise at that time.

---

[10] Van Inwagen (1983: 8) says,
When I say of a man that he 'has free will' I mean that very often, if not always, when he has to choose between two or more mutually incompatible courses of action—that is, courses of action that it is impossible for him to carry out more than one of—each of these courses of action is such that he can, or is able to, or has it within his power to carry it out.

[11] Van Inwagen (1995a: 222) offers an analogy. Dualists and physicalists mean exactly the same thing by the property *being in pain* but offer strikingly different conditions for one's being in pain.

[12] See also Double (1991).

I'm inclined to think van Inwagen is right in saying that there is only one sense of 'could do otherwise' that is operative in debates between compatibilists and incompatibilists. This still leaves room for thinking, however, that there is a sense of 'free will' that includes more than acting while having dual power. Consider Kane (1996a: 15):

> What incompatibilists should claim (and what they have often historically claimed) is that there is *at least one* kind of freedom that is incompatible with determinism, and it is a *significant kind of freedom worth wanting*. They should not quarrel with compatibilists about whether there are freedoms worth wanting that are compatible with determinism, but rather concede the point and go on to argue that compatibilist freedoms are not the only significant ones.

Maybe Kane is right in holding that there are different kinds of freedom that are operative in disputes between compatibilists and incompatibilists. Maybe he is wrong. I propose the more modest claim that perhaps there is a pre-theoretic feature that free will requires that is not itself obviously required by having dual ability, *viz.*, being an ultimate originator.

Even though Honderich urges that we should not be deluded in thinking that there is a single definition of a free choice, he describes at least one sufficiently well-behaved notion of free will, which corresponds what I have called the strict and demanding kind of free will. Honderich (1993: 2) writes,

> Each of us is supposed to have a kind of personal power to originate choices and decisions. Thus on a given occasion, with the past just as it was and the present and ourselves just as they are, we can choose or decide the opposite of what we actually do choose or decide.

I contend that this strict and demanding sense of free will, as Honderich's claim implies, has at least two components. The first one, that acting freely implies dual ability, has already been discussed at length in this section. The next section, §2.2.2, carefully examines what various experts say about a second component, *viz.*, that free activity requires that the agent— presumably in some metaphysically robust way—originates or is a source of an element essential to her action. I aim, then, to present enough textual evidence to underwrite the claim that, intuitively, free will requires dual ability and, perhaps more importantly, origination.[13]

2.2.2   Origination

> *The rationalist's dilemma: either the free act is possible, or it is not—either*
> *the event originates in me or is imposed on me from outside…*
> —Maurice Merleau-Ponty (1962: 442)

We believe that we make a difference, initiating change in the world. We believe that many paradigmatic cases of making a difference are cases of acting freely. Gary Watson (1987:

---

[13] For all I know, there may or may not be some necessary connection between these two concepts. I note them separately for heuristic purposes.

146) thinks that an agent's free action requires that her action be self-determined in the sense that she herself makes the difference as to whether or not she performs that specific action. Timothy O'Connor (1996: 143, 145-6) agrees that we are "self-determining" beings. When we freely make choices for consciously considered reasons, how we act is "up to us." Honderich (1988: 389) observes that we have a pre-philosophical and pre-theoretical "idea or whatever of a determinate centre, a self, which is uncaused in its activity and which is not superfluous." And concerning a free action, Wolf (1990: 10) believes that "…there is a requirement that the agent's control be ultimate—her will must be determined by her self, and her self must not, in turn, be determined by anything external to itself."[14]

Free agents, then, are self-determining entities.[15] Arguably, John Locke and Thomas Hobbes see an intimate connection between free will and this notion of self-determination. Locke (1894: Book 2, Chapter 21, §48) writes,

> …[E]very man is put under a necessity, by his constitution as an intelligent being, to be determined in willing by his own thought and judgment what is best for him to do, else he would be under the determination of some other than himself, which is want of liberty.

The last point Locke makes is crucial: liberty (i.e., free will) implies that the agent's willing not be determined by something "other than himself." Similarly, Hobbes (1969: 47) says, "Liberty is the absence of all the impediments to action that are not contained in the nature and intrinsical quality of the agent."[16] Even Benedict Spinoza (1677: Point 1, Proposition 11) proposes that freedom implies self-determination, saying, "That thing is called 'free' which exists from the necessity of its being, and is determined to act by itself alone."

Philosophers are not always altogether clear about what it means to be self-determining. However, to determine something one*self* suggests that the self, *qua* acting entity, is somehow a source of whatever is determined. The notion of self-determination precludes being determined by something else, i.e., by something other than oneself. Note C.A. Campbell, who agrees with Locke and Hobbes to the extent that a free action is a matter of self-determination. Campbell claims that a free agent is "the *sole* author" and that "the *self* determines" her action.[17] Here, Campbell makes the point that a self-determining agent is a source of whatever is determined by the agent. Consider also Thomas Flint's comments on the relationship between self-determination and freedom. Flint (1998a: 23) states,

---

[14] Wolf (1990) calls this the requirement of autonomy. Her label is fitting, for, etymologically, the term 'autonomy' means self-rule or self-government. See Mele (1995: 3, 162).

[15] See Ekstrom (2000: 3). Galen Strawson (1995: 14) asserts, "…[O]ne is truly self-determining, in one's actions, only if one is truly *self-determined*, and one is truly self-determined if and only if one has somehow or other *determined how one is in such a way that one is truly responsible for how one is*." Of course, someone may be responsible for how one is (in the relevant sense) without being responsible for *every* property that she has. For, consider an agent x that is responsible for how she is and has the property *being such that x began to exist*.

[16] Watson (1987: 146) comments on this quotation from Hobbes, "What is 'intrinsical' to the self cannot be an impediment to the self, and action determined thereby is self-determined."

[17] Campbell (1957: 156-7, 160). See also Campbell (1966: 131-5) and (1967: 48-9).

Self-determination lies at the very heart of freedom; to say that an act of mine was free, but was ultimately determined by someone or something other than me, someone or something whose determining activity was utterly beyond my power to control, is to speak nonsense.

It is reasonable to conclude that, according to these philosophers, acting freely implies that the agent is self-determining, where self-determination implies some kind of origination (yet to be spelled out). Randolph Clarke (2003: 15) distills this point, claiming that when an agent acts with free will she "…determines, herself, what she does; she is an ultimate source or origin or initiator of her behavior." So acting freely, for many philosophers, requires originating occurrences in the world. Plantinga (1990: 31) confirms this line of thought, saying that "the notion of a person as an ultimate source of action" lies at the heart of certain important controversies regarding free will. Van Inwagen (1983: 11) urges that

…the concept of an agent's power to act [freely] would seem not to be the concept of a power that is dispositional or reactive, but rather the concept of a power to *originate* changes in the environment.

Consider also Robert Nozick. Commenting on common beliefs about acting freely, Nozick (1981: 313, *emphasis added*) asserts, "We want it to be true that in that very same situation we could have done (*significantly*) otherwise, *so that* our actions will have *originative* value."
When an agent is an ultimate source or originator of some action, the action is, in as strict a sense as possible, hers—which is precisely what a free action must be. O'Connor (1996: 145) states, "I am quite literally the cause (source, point of origination) of my own behavior." Metaphor permitting, O'Connor (1993a: 500) proposes that free action requires that the agent's behavior is an "outflowing of the agent" in the right sort of way. Thomas Reid (1895: 602) presses, "If the person was the cause of that determination of his own will, he was free in that action, and it is justly imputed to him, whether it be good or bad." Finally, Robert Kane (1996a: 4, 15, 78, 196) defines free will as "the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes."[18]
Clearly, then, many philosophers think that free will requires origination of some sort. An agent is a source of her action when she *herself* ensures which specific action occurs. She is self-determining. Joining this thought to our conclusion drawn from the last section, §2.2.1, we may conclude that free will implies that the agent both has dual power and, upon exercising this dual power, originates change in the world.

2.2.3   The Metaphysical Upshot
The moral gleaned from the previous two sections is this. Free agents have dual power and exercise this dual power in acting freely. If an agent acts freely at a given time, she has it within her power not to perform that action at that time. For example, she could do otherwise at that time. Moreover, a free agent has power to originate change. Free actions demonstrate this

_____

[18] See also Michael Zimmerman (1988: 24-5).

power of origination—in virtue of the fact that in acting freely an agent is, when acting, an underived source of some things that happen in the world. Intuitively, in acting freely the agent herself (*qua* source) makes the difference (via exercising her dual power) as to whether or not she performs that specific action.

How does one specify precisely the intuitive though still imprecisely understood concepts of dual power and origination? What are the exact conditions for an agent's being able not to perform the action she actually performs? What are the exact conditions for an agent's being an ultimate source or point of origination of an action? Neither this section nor the two sections above answer these questions. The point has been only to get a clearer picture of free will by looking at what "the wise and the few" think freedom intuitively involves. We have painted in broad strokes. However, after canvassing the introductory remarks of many philosophers as well as our intuitions, we see that there is ample reason to think that free will requires both dual power and origination.

Exploiting what we've learned from many of the experts, we may buttress what the ordinary folk mean when they say that they possess free will. For example, we may compile what we've learned into a single illustration developing along the following lines.

Free will is making decisions based on what I want to do, without being forced to make a specific decision by something other than myself. I perpetually perform actions, and I act in a wide variety of ways—from singing songs to purchasing pets, from boarding vehicles to scaling cliffs, from planning vacations to eating asparagus. And some of my actions are free. Many of my free actions involve intending to do certain things and following through by actually doing them. I frequently encounter situations where I am conflicted about what to do—I struggle to make a decision. I deliberate about the incompatible things I want, about my beliefs, and (in particular) about what I think would be good for me to do. My deliberation results in making a choice, which in turn leads to my completing the action. I choose from various courses of action each of which I know is well within my power to perform. Each alternative course of action is open to me. I cannot perform them all, as my options are incompatible. So I pick one and go with it. I have the power to go against various factors influencing me. I have the power to go against the grain, against the pressures that vie for one particular action over another. I realize that with respect to some of my actions, I am able to act in a certain way and I am able not to act in that way.

For example, yesterday I deliberated about whether to order either soup and salad or just a sandwich. I wanted the soup and salad. I also wanted the sandwich. But I knew I could only order one. I was able to order the sandwich. Indeed, I opted for the sandwich. However, I might have chosen differently, thereby refraining from ordering the sandwich. Choosing the sandwich was not inevitable. Which way I ordered was not somehow fixed ahead of time. I'm convinced that ordering the sandwich was up to me and therefore open to me. Nothing beyond my power guaranteed that I ordered the sandwich, since *I* was the one determining my behavior at that moment. I understand that I initiated my activity at the close of my deliberations. I am the one who was

ultimately responsible for placing the order. I was able to have refrained at that moment.[19]

Thus we believe and want it to be the case that the future is unfixed and open. Honderich (1988: 392) writes, "That the desire [for an unfixed future], or at the very least the capability of it, is a general fact about us, in an ordinary if not wholly strict sense, is indubitable."

The modality of this openness of the future, i.e., the modality that it *could* go any one of multiple ways, is not merely epistemic. More precisely, it is not ignorance.[20] Saying that the future *could* turn out so-and-so is not equivalent to saying it *may* turn out so-and-so. Roughly, a given future is epistemically possible so long as believing that the future turns out this way is consistent with the rest of one's considered beliefs. For one to say that the future may turn out so-and-so just means that one does not know that the future will not turn out so-and-so. Epistemic modality is a form of ignorance. In contrast, our aspiration for an unfixed future, Honderich (1988: 388) realizes,

> …is not merely for a future that is in practice unpredictable, with nothing said as to it being fixed or unfixed. We are not satisfied with the idea that all the events which constitute our coming lives are in a category with the event which occurred when the ball on the roulette wheel came to rest unpredictably but inevitably on red 7. What we want are futures that are not settled.[21]

A free agent would not reasonably think to herself, "I do not know what I'm going to do, and this why my future is up to me." A free agent may rather think, "I do not yet know what I'm going to do because I have yet to make my choice. I'll know which choice I'm making precisely when I make the choice."

The conditions of dual power and origination are fundamentally metaphysical. Free activity presumably requires that the agent, in some metaphysically robust way, has both the

---

[19] Compare Timothy O'Connor (1995b: 173). He aims to illustrate and characterize our *pre-theoretical* conception of free action, but I find it unlikely that the average intelligent person (the average reader of the *New York Times*, e.g.) would understand it after only one pass through the text.

[20] Pacé Stephen Wolfram (2002: 751), who reasons,

> But so in the end what makes us think that there is freedom in what a [person] does? In practice the main criterion seems to be that we cannot readily make predictions about the behavior of the system. For certainly if we could, then this would show us that the behavior must be determined in a definite way, and so cannot be free. … [A]s soon as the behavior we see becomes more complex we quickly tend to imagine that it must be associated with some kind of underlying freedom. For at least with traditional intuition it has always seemed quite implausible that any real unpredictability could arise in a system that just follows definite underlying rules.

[21] Honderich's insertion of the term 'merely' at the beginning of this quotation is misleading on the grounds that the issue of predictability is a red herring. Perhaps it is possible that an unfixed future be predicted or even known (by someone other than the agent). If so, then someone may consistently long for an unfixed future while refusing to long for a future that is in practice unpredictable. It is far from obvious that predictability implies unavoidability or vice versa.

power to originate change as well as the power not to originate this change. Of course, if one's freely performing action A essentially requires one's originating change x, then one's originating some change y instead of (i.e., rather than) change x suffices for one's not freely performing A.

Kane (2002a: 5) speaks on behalf of the experts we have examined, concluding, "…[W]e believe we have free will when (a) it is 'up to us' what we choose from an array of alternative possibilities and (b) the origin or source of our choices and actions is in us and not in anyone or anything else over which we have no control." That this ubiquitous belief is implicit in the very moments of acting freely suggests that we believe that, unless we are massively deceived, when one acts freely, one knows that one's action essentially involves one's being a source of change. Up until the moment of initiating the change, it is within one's power not to initiate that change.

Benedict Spinoza thinks that we *are* deceived. He (1677: Point II, Proposition 35) writes,

> Men are deceived if they think themselves free, an opinion which consists only in this, that they are conscious of their actions and ignorant of the causes by which they are determined.

I shall assume without argument that Spinoza is mistaken. Of course, were my project to fail woefully, Spinoza's assertion would be confirmed. Assuming that our common beliefs about freedom are veridical simply seems dialectically appropriate.

2.3   On Why Free Will Matters

We now consider various reasons for thinking that free will is significant. Understanding how many valuable features and activities of our daily lives relate to free will brings into relief why some of humankind's best thinkers have been preoccupied with penetrating its nature. Such valuable features and activities of our daily lives include but are not limited to the following: being morally responsible, attaining life-hopes, exercising practical rationality, deliberating, possessing dignity, giving and receiving forgiveness and love, being creators and authors, and being the rightful subject of a host of reactive attitudes.

Providing a thorough treatment of each of these issues would be a worthwhile study in its own right. My task is merely to give my reader a feel for the sorts of reasons for valuing free action. Thus, as we cursorily cover the above features and activities and how they relate to free action, this section both underscores the value of free action and confirms our previous findings that acting freely requires dual power and origination. We continue to consult the experts as well as our bed-rock intuitions.

2.3.1 Moral Responsibility[22]

> *That the moral laws of nature are often transgressed by man, is undeniable.*
> *If the physical laws of nature make his obedience to the moral laws*
> *to be impossible, then he is, in the literal sense,* born under one law,
> bound unto another, *which contradicts every notion*
> *of a righteous government of the world.*
> —Thomas Reid (1895: 337)

Richard Double (1991: 12) writes, "Free will seems, at first blush, to be something without which our moral responsibility for our actions will be jeopardized." Van Inwagen (1993: 185, 1990: 278) points out, "It is a common opinion that free will is required by morality." This common opinion has primarily motivated, and continues to motivate, an interest in probing the depths of the nature of free will. Van Inwagen (1999: 343) says,

> The main interest of the free-will problem, for most philosophers, derived from their belief that moral responsibility was impossible without free will—without the ability to do otherwise. This belief was the main reason most philosophers had for *caring* about free will enough to invest time and ink in a debate about whether anyone had it or what it was compatible with.

In addition to noting this historical fact about philosophers, van Inwagen seats himself in their company. He contends that there is an inseverable connection between responsibility and free will. The inseverable connection, according to van Inwagen (2001: 10), is simply that "…if one is morally responsible for anything, it follows logically that one has had a free choice about something." Rowe (1991b: 237) agrees, boldly asserting,

> If you are morally responsible for your action then you must have played a role in causing your action and the action must have been done freely. I take this claim to be widely accepted, if not self-evident.

Footnoting this assertion, Rowe (1991b: 237) remarks, "We all agree that this claim is true. What we disagree about is the proper account of what it is for an agent to cause her action or to do it freely." Kane (2002b: 407-8), also agreeing with van Inwagen, believes that one is ultimately morally responsible only if one performs some free action at some time.

Dennett (1995: 55) suggests that addressing the problem of free will involves answering the question, "How can we make sense of the intuition that an agent can only be responsible if he could have done otherwise?" Chisholm (1982: 24-5) takes some steps to render this intuition intelligible, advancing the following simple argument.

> Let us consider some deed, or misdeed, that may be attributed to a responsible agent: one man, say, shot another. If the man *was* responsible for what he did, then, I would urge, what was to happen at the time of the shooting was something

---

[22] For sake of brevity, I sometimes supplant 'moral responsibility' with just 'responsibility'.

that was entirely up to the man himself. There was a moment at which it was true, both that he could have fired the shot and also that he could have refrained from firing it. And if this be so, then, even though he did fire it, he could have done something else instead.

When addressing the relationship between responsibility and free will, many philosophers, in keeping with the experts' remarks above, tend to focus most on the relationship between an agent's being responsible and her having dual power. More interesting and less appreciated, in my view, is the fact that some philosophers focus on the link between responsibility and origination.

In the balance of this section, §2.3.1, I consider a few examples from both camps, beginning in §2.3.1.1 with some quotations from those placing more of a focus on dual power. It is fairly well known in philosophical circles that Harry Frankfurt (1969) has cast doubt on the claim that moral responsibility requires dual power. In particular, Frankfurt argues that a person might be morally responsible for performing an action even though she does not have the power to act otherwise. I outline his case in §2.3.1.2. My response to Frankfurt is two-fold. First (§2.3.1.3), I underscore how Frankfurt's project confirms the claim made by other wise philosophers, *viz.*, that responsibility requires origination. Second (§2.3.1.4), I argue that Frankfurt's project fails to undermine the claim that moral responsibility requires one's acting with dual power at least on some occasion.

2.3.1.1   Moral Responsibility and Dual Power

Here are some more excerpts from experts. Aristotle states that "…if to act, where this is noble, is in our power, not to act, which will be base, will also be in our power, and if not to act, where this is noble, is in our power, to act, which will be base, will also be in our power."[23] Plantinga (1990: 30) asks rhetorically,

> But how can I be responsible for my actions if it was never within my power to perform any action I didn't in fact perform and never within my power to refrain from performing any I did perform?

Ginet (1997: 85) asserts "…that an agent can be morally responsible for her action only if it is a free action: an agent can merit credit or blame for something she did only if she could have done otherwise."[24] Stewart Goetz (1988: 306) holds that "…moral responsibility presupposes the explaining of free human actions in terms of reasons." Since an explanation of free human actions in terms of reasons implies that there are some free actions, it follows that Goetz is committed to the claim that moral responsibility presupposes that there are some free actions. Van Inwagen argues that dual ability is implied by any correct judgment of the form 'You should not have done X,' which counts as one of the most poignant examples of moral responsibility. He (1993: 185) states,

---

[23] Aristotle (*NE* III 5, 1113b7).

[24] Ginet (1996) argues for this claim.

The judgment that you shouldn't have done X implies that you should have done something else instead; that you should have done something else instead implies that there was something else for you to do; that there was something else for you to do implies that you *could* have done something else…

Thomas Reid (1895: 621) also claims that responsibility requires dual ability. He states,

Another thing implied in the notion of a moral and accountable being, is power to do what he is accountable for. That no man can be under a moral obligation to do what it is impossible for him to do, or to forbear what it is impossible for him to forbear, is an axiom as self-evident as any in mathematics. It cannot be contradicted, without overturning all notion of moral obligation; nor can there be an exception to it, when it is rightly understood.

At first glance, this does not seem to commit Reid to holding that an agent performing an action for which she is accountable had it within her power *not* to perform that action; the quotation above suggests that responsibility for X requires power to bring about X. Nonetheless, lest we forget what we learned about Reid's views in §2.2.1, recall that he clearly holds that power to act implies power not to so act. Reid (1895: 65) states, "Power to produce an effect supposes power not to produce it; otherwise it is not power but necessity, which is incompatible with power taken in the strict sense."[25] So, Reid thinks that being morally obligated to perform some action implies power to perform that action, and since he believes that power to perform a given action implies power not to perform that action, Reid holds that moral obligation requires dual ability.

Thus, many philosophers believe that responsibility requires dual power. Nonetheless, Harry Frankfurt's (1969) seminal essay 'Alternate Possibilities and Moral Responsibility' raises considerable skepticism against this common belief. I ultimately find Frankfurt's case unconvincing. The next section, §2.3.1.2, merely sketches the gist of Frankfurt's argument. Then, before criticizing Frankfurt's argument in §2.3.1.4, I discuss in §2.3.1.3 how Frankfurt's project nicely supports the relatively popular claim that responsibility requires origination. While one may concede quite a bit to Frankfurt, I argue that his project confirms the claim that moral responsibility requires acting with dual power on some occasion. I shall conclude that one is quite reasonable in thinking that moral responsibility implies that there is a free action. And since moral responsibility is important, so too is freedom.

---

[25] Thus Reid reiterates Aristotle's (*NE* III 5, 1113b7-8) dictum, "For where it is in our power to act it is also in our power not to act, and *vice versa*."

2.3.1.2   Harry Frankfurt[26]

This section briefly outlines the upshot of Frankfurt's project. Frankfurt aims to show that (i) the freedom pertinent to moral responsibility is an agent's acting of his own accord, and (ii) this freedom does not strictly imply that the agent has the power to act otherwise. Hence, Frankfurt contests the *Principle of Alternative Possibilities* (PAP), which states,

(PAP) A person is morally responsible for performing a given act A only if he could have acted otherwise.

Frankfurt's argument against PAP consists of two steps. Frankfurt argues that there could be background circumstances ensuring that someone could *not* have done otherwise but where these background circumstances do not in any way influence the action she performs. Since the agent could not have done otherwise, the situation in which the agent acts is an *irresistible-situation* (*IRR-situation*, for short). Frankfurt contends that, intuitively, someone might still be morally responsible for performing an action in an IRR-situation. If Frankfurt is right, then it follows that PAP is not necessarily true, i.e., possibly, someone is responsible for performing an action A even though she has no alternative but to perform A. One may thus reasonably hold that responsibility need not require dual power.

The following thought-experiment captures the essentials of Frankfurt's counterexample to PAP—it is an attempt to articulate an IRR-situation:

> Jones is in the position to inherit a large sum of money from a distant relative. However, the sum would be even greater were he not to have to split the total inheritance with his nephew, Smith. Jones is torn over whether or not to kill Smith. Jones obviously has reasons to kill him, terrible as they may be. But Jones is not an altogether mean man. He thus also has reasons against killing him. They vacation together in Arizona, touring the south rim of the Grand Canyon. After more contemplation, Jones greedily chooses to kill Smith in order to secure a larger inheritance. Jones succeeds in killing Smith by pushing him off the rim of the canyon. Many take it as obvious that Jones is morally responsible for killing Smith. Intuitively, Jones is responsible for Smith's death.
>
> However, this is not the entire story. In addition to these circumstance and unbeknownst to Jones, Black, a very powerful demon, also wants Smith out of the picture. Black goes to great lengths to ensure that Smith dies and would prefer that Jones do the dirty work. So, Black monitors every nuance of Jones's behavior and mentality. If Black sees that Jones will not choose to kill Smith (Black is an excellent

---

[26] The corpus of work following Frankfurt's (1969) seminal article is dauntingly huge. E.g., see Robert Allen (1997); David Copp (1997); John Martin Fischer (1999: 98 note 8, 1995, 1994: Chapter 7, 1986a, 1982); Fischer & Mark Ravizza (1998); Robert Heinaman (1986); James Lamb (1993); Michael McKenna (1997); Alfred Mele (1995: 65-80), Mele and David Robb (1998); Margery Naylor (1984); Timothy O'Connor (2000a: 18ff); Michael Otsuka (1998); Derk Pereboom (2000, 1996); Gordon Pettit (2000); William Rowe (1991a, 1989); Eleanore Stump (1996, 1995); van Inwagen (2001: 10-14, 1986b: 155ff, 1983: 167-171, 1978); Widerker (2000, 1995a, 1995b, 1991); Keith Wyma (1997); and Linda Zagzebski (2000).

judge of such things), Black will intervene and directly cause Jones to choose to kill Smith and to carry out this implanted intention by pushing Smith off the cliff.[27] Black, however, prefers not to show his hand unnecessarily. So if Black sees that Jones will choose to kill Smith on his own, Black does nothing. Either way Jones should choose, Black gets what he wants. As things turn out, though, Black does nothing since Jones chooses and kills Smith on his own. Black's committed presence is causally innocuous and therefore irrelevant to the actual production of Jones's behavior.

Call the above situation in which Jones acts the *Black-situation* (i.e., *B-situation*). Let 'A' refer to Jones's act of making a choice to kill Smith.[28]

Roughly, Frankfurt believes that the B-situation is an IRR-situation, since the B-situation implies that Jones has no ability not to perform A (due to Black's scheming). Intuitively, Jones is morally responsible for A whether or not Black is in the offing, ready to intervene manipulatively. But since Black lurks only in the shadows, Jones does not have dual power with respect to A. Or so the intuition goes. Thus, responsibility seems not to imply dual power, though admittedly the two are often instanced together. Frankfurt concludes that since Jones is responsible for performing A but performs A without dual ability, responsibility does not require dual ability.

2.3.1.3 Moral Responsibility and Origination: Frankfurt and More Wise Opinions Confirm It

While Frankfurt's argument raises skepticism against the traditionally accepted claim that responsibility requires dual power, I want to bracket the question of whether or not his argument succeeds. I want now to underscore a different point that Frankfurt earns along the way. More precisely, I discuss how Frankfurt's argument succeeds at one level, illustrating my contention that responsibility requires origination. Thus, if some criticism against Frankfurt's ultimate conclusion is correct (as I shall argue in the next section), then my study vindicates the claim that responsibility requires acting freely. I finish this section by mentioning what other philosophers say about the relationship between responsibility and origination.

Recall that intuitions dictate that Jones is (actually) responsible for performing A, where A is his choosing to kill Smith. Were Black to intervene, however, Jones would not be responsible for performing A. So, *why think Jones is actually responsible but would not be responsible were Black to intervene?*

One difference between the actual case and the counterfactual case is almost too obvious to be worth stating. In the actual situation, Jones performs the act of willing (i.e., choosing) to kill Smith on his own. In the counterfactual case, Jones does not will to kill Smith on his own, for Black directly causes Jones to will to kill Smith. Smith would act, yes. But Black would work through Smith too. Therefore, it is reasonable to suppose that we're convinced that Jones in the actual case is responsible because *he himself* wills, *he himself* chooses. Jones wills to kill Smith by himself. He acts alone. We think Jones's choosing to kill Smith involves some sort of origination on the part of Jones. Were Black to intervene, Jones would not be responsible

---

[27] I'm assuming, here, that it is possible that Black directly causes Jones to choose to kill Smith.

[28] Choosing to do something is itself a doing. Making a choice is doing something. Making a choice may not be an overt action, but it is an action nonetheless.

because Black, not Jones, would be initiating Jones's act of choosing to kill Smith. That Jones actually initiates the choice/willing but initiates nothing in the counterfactual case (since Black, not Jones, initiates the action) helps explain the difference in moral status between the actual case and the counterfactual case.

Thus, the Frankfurt case confirms the claim that responsibility implies origination. Recall that one of Frankfurt's principal aims is to establish that the freedom pertinent to responsibility is an agent's acting of his own accord. I think that Frankfurt earns this much.

To preempt a misunderstanding and in order to get a better handle on the concept of origination, we should be careful not to confuse *possession* with origination. For, someone can perform an act that is *his* without originating anything. Thus, to say that Black directly causes Jones to will to kill Smith does *not* imply that this act of willing is not Jones's. If Black causes Jones to will to kill Smith, the act of willing is still Jones's. Such possession, though, does not require origination. Rowe (1991b: 254) correctly notes that

> …regardless of causal origin, the volition will be Jones's act of willing. For it is conceptually impossible for a volition to kill Smith to occur in Jones without it being true that *Jones* wills to kill Smith. Willing (like thinking, imagining, and believing) cannot occur in a person without it being true that *the person* wills (thinks, imagines, believes). So even if Jones does not cause the volition to kill Smith, Jones, nevertheless, wills to kill Smith and performs the action of killing him.[29]

So, even though Jones performs the act of willing in the event that Black causes this volition, Jones's choosing does not imply that he originates anything in the world.

Many philosophers think—independently of their views on Frankfurt's argument—that responsibility implies origination. For example, predating Frankfurt by over two millennia, Aristotle (*EE* II 6, 1223a9-15, *emphasis added*) asserts in his *Eudemian Ethics*:

> And since excellence and badness and the acts that spring from them are respectively praised or blamed—for we do not give praise or blame for what is due to necessity, or chance, or nature, but *only for what we ourselves are causes of*; for what another is the cause of, for that he bears the blame or praise—it is clear that excellence and badness have to do with matters *where the man himself is the cause and source of his acts*.

Aristotle's remarks nicely recapitulate the conclusion drawn from the Frankfurt scenario in the previous section.

More recently, Derk Pereboom (2001: 10) thinks that

> …for an agent to be morally responsible for an action is just for the action *really to belong to the agent*. Equivalently, but in classical phrasing, for an agent to be morally responsible for an action is for it to be *imputable* to the agent.

---

[29] I've innocuously adjusted Rowe's quotation to refer to the characters in the preceding example.

Intuitively, the difference between *really* belonging to the agent and belonging to the agent *simpliciter* lies in the distinction (outlined in the previous section) between possession and origination. Honderich (1996: 858) puts the point a bit more crisply than Pereboom, saying that "…origination has to be a fact if we are to have everything we want in connection with freedom, responsibility, and so on." And consider John Bishop (1989: 1), who states that responsibility requires "the 'originative' ability to initiate events in the natural world."

Regarding the intimate nexus between responsibility and the condition of origination, Wolf (1990: 10-11, *my emphasis*) writes:

> This condition, like others, seems to cohere with the meaning responsibility has for us. It makes sense that beings who can purposefully *initiate* change should have a different status in the world from that of those who merely execute it. It makes sense that such beings should have a special significance, for they are *sources* of value (and disvalue) rather than mere carriers of it.

Finally, Fischer considers two ideas. The first idea is that there are various, genuinely available paths to an agent. Having various paths that are genuinely available corresponds to what I have called dual ability. The second idea is that the agent herself selects the path rather than some other force or chance. Saying that the agent herself selects the path rather than some other force or chance corresponds to the agent being a source of her activity. Fischer (1999: 99, *my emphasis*) asserts,

> It seems to me that *both ideas* are important components of the traditional conception of the sort of control associated with moral responsibility—alternative-possibilities control.

Thus, as Fischer construes the traditional conception of what moral responsibility requires, responsibility requires dual ability and origination.

It appears, then, that moral responsibility and origination are bedfellows. Intuitively, the former requires the latter. Even Frankfurt's project confirms this result. Let's turn, then, to see whether or not Frankfurt undermines the traditional belief that responsibility requires dual power.

2.3.1.4   Critically Evaluating Frankfurt's Case

I aim to show that Frankfurt-style criticisms fail to undermine a principle implying that moral responsibility requires dual power, whether that principle is PAP or some other principle in the neighborhood of PAP. I rehearse van Inwagen's argument for thinking that PAP is false only to replace PAP with a similar principle that I argue is immune to Frankfurt-style criticisms. If I am right, Frankfurt-style considerations provide no ground for thinking that one can act responsibly without ever acting with dual ability.

To review,

(PAP) A person is morally responsible for performing a given act A only if he could have acted otherwise.

Van Inwagen (2001: 12) claims that PAP is false, if not meaningless. On whether PAP is meaningful, van Inwagen reports, "It has always seemed reasonably plain to [me] that what one is morally responsible for is not one's actions but the <u>consequences</u> of one's actions, or, more exactly, <u>certain</u> of the consequences of one's actions." One might ask, "To what does the term 'consequences' refer?" A broadly logical consequence of *one's performing a specific action A* is *one's performing A*. Apparently, this is not what van Inwagen has in mind. Van Inwagen (2001: 12) asserts,

> …[W]hatever ontological category one thinks the consequences of a person's actions should be assigned to, one should assign them to the same ontological category or categories as the consequences of an earthquake or a scientific discovery or a rise in the prime lending rate.

Common parlance confirms van Inwagen's intuition. For example, we normally do not blame a doctor for making an incision with the scalpel. Rather, we usually blame doctors for what results from their actions, such as the patient's dying. Intuitively, the only reason one would blame the doctor for making the incision is *because* the patient died. Van Inwagen (2001: 12) says, "[I have] never been able to convince [myself] that [I] understand sentences like 'Bill is morally responsible for lying under oath'." He concludes that since one cannot be responsible for any of one's actions, then PAP expresses no proposition whatsoever. It is meaningless.

One might still wonder: why not think that at least one of the consequences of an earthquake is the earthquake itself? After all, some consequences are logical consequences. So why think a doctor could not be responsible for *both* his making the incision as well as the patient's dying? Being responsible for making the incision may depend on being responsible for one of its consequences, say, the patient's dying. The responsibility for the former, then, would be derivative. But it would still be responsibility all the same. I do not find it obvious that responsibility for one's action could not be derivative in this way. And it seems reasonable to think that it could be derivative. Moreover, perhaps someone's action *A is* a (causal and temporal) consequence of some previous action *B*? Hence, since *A* is a consequence—even by van Inwagen's lights—it is the right sort of thing for which one may be responsible.

Regardless of whether one thinks that van Inwagen's objection is telling, fixing PAP is easy. Just let it read, "Someone is morally responsible either for performing action A or for some of the action's consequences only if he could have acted otherwise."

Supposing, for sake of argument, that PAP is meaningful, van Inwagen concedes that it is false. What is more, one can see PAP's falsity without elaborate and contrived scenarios such as Frankfurt-style examples. For example, van Inwagen considers a hardened thug who kills his hundredth victim. The thug is responsible for the victim's death, but he was not then able to refrain from pulling the trigger. Thus, PAP is false, strictly speaking. Examples like this one are not hard to find. The seasoned alcoholic who drives inebriatedly and hits a pedestrian is responsible, though he was unable to swerve her car. Or, suppose that I owe you a large sum of money. I promise to repay you at Vincent's Café next Monday at noon. I never show up. In your thoughts, you begin to blame me for being somewhere else at the appointed time. Your disgust with me, however, temporarily subsides when you learn that I was locked inside a bank vault on Monday. At noon, I could not repay you at Vincent's Café. I could not but be somewhere else. But then you learn that I locked myself in the vault that Monday morning in

order to avoid repaying you. Intuitively, I'm responsible for being somewhere else at noon even though at noon I could not be at Vincent's Café. Thus, PAP is false, strictly speaking.

Intuitively, the hardened thug's responsibility derives from previous actions he performs (or from consequences of those previous actions). He is responsible for becoming the sort of person who cannot refrain from killing at the order of the don. He is responsible for the hardening of his character. The same goes for the alcoholic and me. I cannot do other than be somewhere other than Vincent's Café *once* I'm locked inside the vault. Nonetheless, my responsibility for being somewhere else derives from being responsible for locking myself in the vault.

Following Aristotle (*NE* III 5, 1114a13-22), van Inwagen (2001: 8) concludes that "…although there is an inseverable connection between free will and moral responsibility, this connection, inseverable though it be, can be stretched exceeding fine."[30] Amending PAP accordingly, we get

(PAP*)　　　A person is morally responsible for performing action *A* or for some of *A*'s consequences only if, at some time no later than *A*, he could have not performed some action *A** that he actually performed.

Does Frankfurt's objection apply equally well against PAP*? I think not. For, I shall argue that Frankfurt cannot describe an IRR-situation for an agent performing her first morally significant action. Say that some agent Δ's action A is *significant* if and only if either Δ is morally responsible for performing A or Δ is morally responsible for some consequence of *Δ's performing A*. Say that an agent Δ's action A is her *first* action if and only if Δ performs A, and Δ performs no other action A* prior to A. If an IRR-situation is not one in which the agent performs her *first* morally significant action, the one may reasonably claim that the responsibility in the IRR-situation derives from some other significant action the agent performs. So the crucial question is whether there could be an IRR-situation for an agent's first significant action. But I shall argue that there is no IRR-situation in which an agent performs her first significant action. Therefore, despite Frankfurt's efforts, one may reasonably hold that moral responsibility requires some agent's acting freely on some occasion.

Consider an agent performing his first significant action A. I think that most incompatibilists believe that A will be an undertaking, an endeavoring, a choosing, a decision, or the like. The rationale can be traced back to Aristotle, who (*EE* II, 1228a10-17) asserts,

> Further, we praise and blame all men with regard to their choice rather than their acts (though activity is more desirable than excellence), because men may do bad acts under compulsion, but no one chooses them under compulsion. Further, it is only because it is not easy to see the nature of a man's choice that we are forced

---

[30] Van Inwagen (2001: 10) says, "The inseverable connection is this: if one is morally responsible for anything, it follows logically that one has had a free choice about something." Van Inwagen (2001: 10-14) advances his *principle of possible prevention*, which basically states, "A person is morally responsible for a certain state of affairs only if that state of affairs obtains and there was a time at which he could so have acted that that state of affairs not obtain." See van Inwagen (1986b: 155ff, 1983: 167-171) for an earlier, negligibly different, version of this principle.

to judge of his character by his actions. The activity then is more desirable, but the choice more praiseworthy.

Fortunately, there's little we need to change to my Frankfurt-style example discussed above. Jones chooses to kill Smith on his own. Black sits in the shadows, ready to make Jones choose to kill Smith were he to see that Jones not choose to kill Smith on his own. We need only to *add* one feature to the example. We stipulate that making the choice to kill Smith is the first significant action Jones performs.[31] So, let the B-situation include this stipulation as well.

Recall that 'A' refers to Jones's act of choosing to kill Smith. Recall also that the B-situation includes Jones's performing A on his own. Let's grant that the B-situation also implies

(1)   Were Jones not to perform A on his own (at time T), Jones would still perform A (at time T).[32]

Proposition (1) lies at the heart of Frankfurt-style examples.   The key is that, intuitively, (1) strictly implies that Jones could not do other than perform A. That is,

(2)   Necessarily, (1) is true only if Jones could not do otherwise.

The reasoning behind (2) is simple: regardless of what Jones does on his own, Jones will still do A. Now, Frankfurt may argue as follows. The B-situation implies both that Jones performs A on his own and that (1) is true. From this result and (2), it follows that

(3)   Jones could not do otherwise.

Thus, the B-situation is an IRR-situation. Hence, it is possible that someone performs his first significant action without dual power. Hence, it is possible that someone is morally responsible for performing an action without dual power. Therefore, moral responsibility does not require any dual power whatsoever.

As I see it, one problem with this argument resides in the opacity of (2). Suppose, for sake of argument, that Frankfurt earns the truth of (1). One may still wonder about whether or not the conclusion is relevant to whether or not Jones performs A on his own with dual power. Frankfurt would contend that (3) bears directly on whether or not Jones performs A on his own with dual power, since having no ability to do otherwise is equivalent to having no dual power.

I argue that the incompatibilist may reasonably think that (3) is opaque regarding precisely *what* Jones cannot but do. To make (3) more precise, one needs to fill the ellipses in

(3*)   Jones could not do other than… .

Moreover, if (3) is opaque, (2) inherits the opacity as well, as (3) just is the consequent of (2). So, Frankfurt's criticism goes through only if the appropriate version of (2) is relevant to whether

---

[31] If this stipulation stretches one's intuitions too much, one may alter the example accordingly.

[32] Hereafter, I drop but still assume the temporal indices.

or not Jones performs A on his own with dual ability, i.e., only if filling in the ellipses in the proposition

(2*)   Necessarily, (1) is true only if Jones could not do other than…

is relevant to whether Jones performs A on his own with dual ability. To simplify the discussion, I focus on (3*). My points apply *mutatis mutandis* for (2*).
          Here is one obvious candidate for (3*):

(3**)   Jones could not do other than perform A.

          I think that Frankfurt has (3**) in mind. I grant that if Frankfurt earns (3**), then Jones does not have dual ability *with respect to* performing A. However, (3**) need not undermine the claim that responsibility requires dual ability. For Jones may have dual ability with respect to the obtaining of a different state of affairs. The object of dual ability (i.e., that over which one has dual ability) may be more finely-grained than initially supposed. Thus, even if (3**) implies that Jones has no dual ability with respect to performing A, Jones may still have dual ability with respect to some other state of affairs obtaining, and Jones's responsibility may require this more finely-grained dual ability.
          One may reasonably think that Jones's dual ability lies in the fact that Jones has the power *not* to perform A *freely*. For Jones has the power not to do A freely (at least partly) in virtue of having the power not to do A on his own, not to perform A on his own accord, not to originate change in the world, not to be a source of change in the world. I contend that Jones acts with dual ability in virtue of being able not to perform A on his own accord.
          So notice that Frankfurt does nothing to show the falsity of

(4)   Necessarily, if Jones performs A on his own, then Jones has the power not to perform A on his own.

The B-situation, if self-coherent, is consistent with Jones having the power not to perform A on his own. In short, the B-situation confirms the truth of (4), and (4) together with the proposition that Jones performs A on his own implies that Jones acts with dual ability. Hence, the B-situation implies that Jones acts on his own with dual ability. Surprisingly, Frankfurt's project confirms the claim that responsibility requires dual power.
          Here's another way to illustrate the same basic point. Black, the counterfactual manipulator, does not actually intervene. By hypothesis, A is significant (from a moral point of view), and Jones performs A on his own. Now, consider a nearby scenario where Black intervenes, directly making Jones perform A. Why does Black intervene? Because he *sees* that Jones will not perform A on his own. 'Sees' is a success term. What is seen is there to be seen. Thus, 'sees that' implies 'it is true that'. This is important because the proposition expressing what Black would see in the counterfactual situation at least partly explains the actual dual ability of Jones. I conclude that Jones could refrain from doing something—we just need to be more careful about what precisely it is over which he has power to refrain from doing.
          I contend, then, that one *need* not hold that Jones could refrain from making the choice to kill Smith but may instead hold that he could refrain from making the choice to kill Smith freely (or, on his own). Thus, our discussion confirms the claim that Jones performs A on his own, and

Jones has the power not to perform A on his own (as illustrated by the content of what Black would see were he to intervene). In fine, the truth of (4) may capture the intuitive claim that Jones responsibly performs A on his own and with dual ability.

        If I am right, then we may fill in the lacuna of (3*) as follows.

(3***)   Jones could not do other than perform A on his own.

Consequently, the corresponding instance of (2*) reads,

(2**)   Necessarily, if Jones would still perform A were he not to perform A on his own, then Jones could not do other than perform A on his own.

But as we have seen above, Frankfurt-style cases provide no good reason for thinking that (2**) is true.

        In conclusion, there are good reasons to think that Frankfurt-style considerations do not show that an agent can perform her first significant action without dual power. A finite agent performs a responsible action only if she performs some first significant action. Frankfurt's thought-experiment fails to undermine the claim that an agent performs her first significant action only if she acts with dual power. Hence, a proponent of Frankfurt-style examples fails to undermine PAP*. Thus, the intuition that moral responsibility requires origination and dual power is vindicated.

        It appears that moral responsibility requires a certain kind of free will, namely, that kind implying both dual power (as §2.3.1.1 and this section, §2.3.1.4, suggests) and origination (as §2.3.1.3 suggests). If this is right, then let this kind of free will be the operative notion in our study. Thus, since moral responsibility is something we obviously care about, we have ample reason to care about free action. Since moral responsibility matters, freedom matters.

2.3.2   Attainable Life-Hopes

        There are other features of our lives besides being morally responsible that are connected with freedom. One of them is the ability to achieve our long term goals, i.e., our life-hopes. From time to time we dream about our futures. We set goals, make resolutions and promises, and hope that things will turn out as we plan. We form values about what is important to us, and these values guide our general activities. We trust that we are to some degree in control of how well these values guide our behavior and therefore of how close we come to attaining our hopes and goals. We generally assume that we will often be in the position where it is both the case that we are able to act in such a way so as better to achieve our goals and the case that we are able to act in such a way so as to frustrate achieving these goals. We believe that in such moments we will be sources of our wise or foolish decisions. In short, we believe that our practical rationality implies our target sense of freedom.

        For example, suppose that in an unprecedented moment of deep reflection a teenager commits to herself to earn a college degree and become a commercial airline pilot. She knows that there are a host of obstacles that may crop up and squash her hopes. The obstacles may be so devastating that she simply cannot become a pilot—e.g., perhaps she will be blinded accidentally or die prematurely. The obstacles, however, may come in various forms of deterrents, which may give her reason to quit seeking out her hope of piloting an aircraft.

Nonetheless, she believes that such deterrents may be overcome by initiating new intentions and actions. "I've have the will power," she whispers to herself. The teenager, reasonably enough, believes that she will later be in the position to surmount many, and hopefully enough, of these unforeseen hurdles by responding to them appropriately. She believes that her accommodating actions will not just happen to her. She will not merely react. But they will instead be her own mark on the world. She will respond. She commits in her heart to respond wisely, even though she foresees the inevitable likelihood that she will be tempted to act unwisely on specious desires some of which will likely be inherited down the road. Then, breathing a small sigh of satisfaction, she opts against going to the cinema with her friends and gets head start on a weekend's worth of pre-calculus homework.

Honderich (1988: 384) writes that we, like the teenager,

> …somehow believe or conjecture that *we stand to our actions in such an initiating way that we have at least some chance of fulfilling our hopes*. This initiation is integral to the idea of achievement, and as relevant…to actions as means. …It is (I say) I who will act, and so may come to have what I want. It is I who will give rise to my actions, or bring them about, and so can have some optimism.

So it seems to us now that we will have a choice about how we respond to unforeseen factors that come between us and our life-hopes. We will then be in a position to act wisely or foolishly. On the majority of these occasions, we will be frequently disposed to respond in certain ways. However, Honderich (1988: 386) charges, "I am not inevitably the creature of my dispositions, but rather I have the chance—even if only a chance—of being their master." Presumably Honderich speaks not only for himself but for nearly everyone else as well.

It seems, then, that we have practical rationality. That is, we have the power either to give up on our goals and hopes or to press on, adjusting our behavior so as to once again, at least for all we know, be in a position to secure our goals. This is a conviction the healthy among us share. We believe that somehow we as agents have a final say about what sorts of people we become, for good or for bad.

There is reason for believing that, in normal circumstances, attaining one's life-hopes presumes our strict and demanding freedom. If the freedom required by the attainability of life-hopes is the same freedom moral responsibility requires, then let this be the operative notion in our study. Now, given the importance of our life-hopes, we should value our free will. Since it matters that we are able (and not able) to adjust our behavior to secure our life-hopes, free will matters.[33]

---

[33] For more on why practical rationality (e.g., adjusting means to ends) requires free will, see Reid (1895: Essay 4, Chapter 8). There is an obvious analogy between acting morally (addressed in the last section) and acting wisely (addressed in this section). Both cases are normative, broadly speaking. The former is morally normative, and the latter, prudentially normative. However, since it seems to me that some cases of prudential action are not always morally significant, I dedicate a section for each.

2.3.3 Deliberation

No man can deliberately attempt what he does not believe to be in his power.
—Thomas Reid (1895: 269)

Our deliberating about what we should do supports our conviction that we are free. We frequently deliberate about what we should think, what we should intend to do, what we should say, when we should leave, or how we should solve a problem. What would life be like without any deliberation? It is hard to imagine.

For example, imagine playing a game of chess without any deliberation, either during your turn or during your opponent's turn. Or imagine what it would be like to be a jury member were you unable to deliberate. To the extent that we believe that we have free will, I think that our free will would be severely diminished without deliberation. For, we believe that in the typical case of deliberation, the person ends her deliberation precisely when she acts freely. We value deliberation, as we find it arduous even to imagine our lives any better without it. Indeed, we have a strong suspicion that our lives would be much worse without some deliberation. Initially, it seems that since deliberation matters, so too free will matters.

Richard Taylor (1992: 39-40) notes that when we deliberate we seem to presume several things, regardless of whether we think about them explicitly. Taylor holds (correctly, I think) that one's deliberating *prima facie* implies one's conviction in free will.[34] Since the balance of my book makes reference to deliberation, discussing each Taylor's four pre-theoretic features of deliberation benefits our study.

First, Taylor notes that one can deliberate only about one's own behavior and never about the behavior of another. We can guess or speculate about what someone else will do. We can imagine what another person will do in response to what we do at the close of our deliberations. But we cannot deliberate about what another person will do.

Second, one deliberates only about future things, never things past or present. Of course, we can speculate about what occurred in the past or about what is occurring now. We can guess what we might have done. We can try to remember what we did. But we do not deliberate about past events. Similarly, if I'm now riding a bicycle I can't rationally deliberate about whether or not I'm now riding a bicycle. Of course, I can contemplate whether I'm now riding a bicycle. But mere contemplation is not deliberation. I may deliberate about how much longer to ride the bicycle. But this is about the future. Deliberation, then, concerns what is yet to be done.

Third, one cannot deliberate about what one shall do if one already knows what one is going to do. Suppose that I am in a room and am required to leave. I must choose a door through which I will leave the room. Do I walk through the green door, or do I walk through the red door? Suppose, though, that I also know that regardless of how much I think about it, a demon will make me walk through the green door. It seems rather obvious that I cannot then deliberate about whether or not I will walk through the green door. For I already know what I'll do. The issue is settled in my mind, and so there is no deliberating about which door I will walk through.

---

[34] See also van Inwagen (1983: 205), who says, "It is because the proposition that we have free will is inseparably bound up with our deliberative life, in my view, that most of us are certain we have free will."

Finally, and most pertinent to our discussion, Taylor notes that someone's deliberating about what to do, even if she may not know what she is going to do, implies that she believes that it is up to her what she is going to do. That is, someone cannot deliberate about what to do, unless she believes that it is up to her what she will do. Imagine, again, the case of the two colored doors. Suppose, though, that instead of knowing that a demon will make me walk through the green door, I simply know that a demon will make me walk through one or other of the doors. So I do not know that he'll make me walk through the green door, and I do not know that he'll make me walk through the red door. Even though I do not know what I shall do, I still cannot deliberate about which door I will walk through. I can speculate. I can wait and see what I'll do. But since I know that it is not up to me to walk through the green door and that it is not up to me to walk through the red door, then I cannot deliberate about what to do.[35]

In fine, we often find ourselves thinking about what we will do in the near, or even distant, future. We deliberate about which of a few competing courses of action we might take. Sometimes we think to ourselves, "Should I do so-and-so, or should I instead do such-and-such?" It seems that we cannot at the same time consistently think to ourselves, "I'm neither free to do so-and-so nor free to do such-and-such." Intuitively, then, one's deliberating about which action to perform presupposes a conviction that one is free to act.

Taylor (1992: 53) concludes:

> Our data—to the effect that we do sometimes deliberate before acting, and that, when we do, we presuppose among other things that it is up to us what we are going to do—rest upon nothing more than fairly common consent. These data might simply be illusions. It might, in fact, be that no one ever deliberates but only imagines that he does, that from pure conceit he supposes himself to be the master of his behavior and the author of his acts.

My aim, here, is to register more of our intuitive beliefs about why freedom is significant. Intuitively, deliberation culminating in action intuitively requires the conviction that one is free. Since deliberation matters, freedom matters.[36]

2.3.4   Dignity

Imagine the following. You have just performed what you consider to be a rather distinguished action. You're convinced that you freely performed the action.   In the immediate wake of your action God speaks to you. You know that God is speaking to you. Finally, God tells you that the action you took to be free really was not. Your action was not free because you played no relevant role in originating the action. Rather, other forces worked through you, causing you to perform the action in question.

---

[35] Van Inwagen (1983: 155) asserts,
In my view, if someone deliberates about whether to do A or to do B, it follows that his behaviour manifests a belief that it is *possible* for him to do A—that he *can* do A, that he has it within his power to do A—and a belief that it is possible for him to do B.

[36] Consider also van Inwagen (1983: 205), who says, "It is because the proposition that we have free will is inseparably bound up with our deliberative life, in my view, that most of us are certain we have free will."

What is the appropriate response, from an epistemic point of view? Perhaps you should first believe yourself fortunate and honored to be addressed directly by God. But what is the appropriate cognitive response with respect to your earlier beliefs about your action? Intuitively, if you are at all typical, you would no longer believe that your action was as distinguished and remarkable as you initially thought. There is a sense that your dignity has been compromised, a sense that does not issue merely from your discovering that you held some false beliefs. Your action no longer seems as impressive. It no longer seems so worthy of respect. You were unwittingly along for the ride, a conduit through which forces other than yourself issued in your action. Richard Double (1991: 12) writes,

> It is clear, though, that free will has to do with making choices that have the desirable property of being *free*, which enables agents who make such choices to be more worthy of dignity than agents who cannot.

Were your action free, you would have been, in the most metaphysically robust sense possible, its source. In being an origin of the action, the action would be, in the strictest and most literal way, rightly attributed to you. Clarke (2003: 7) says,

> It is dignifying, we may reasonably hold, to have events so attributable to oneself, and the freedom of the will that is the basis of this attributability may for this reason be held to be a good thing.

There is dignity in being an originator of your actions, and lack of origination corresponds to diminished dignity. There is a sense of freedom, then, that implies some level of dignity. If this sense of dignifying free will is the same sense of free will required by non-illusory deliberation, the attainability of life-hopes, and moral responsibility, then let this sense be the operative notion in the rest of our study.

2.3.5   Forgiveness
    In the film *Bruce Almighty*, Bruce Nolan (played by Jim Carrey) complains to God about how miserable he thinks his life has become, accusing God of picking on the less fortunate and otherwise neglecting his cosmic duties. God decides to teach Bruce that running the universe is no easy task. God (played by Morgan Freeman) meets with Bruce and endows Bruce with divine powers to govern his hometown in the manner he deems fit. God makes only two stipulations: Bruce can't tell anyone else about his divine powers, and he can't "mess with free will."
    Predictably, Bruce doesn't do so well. He exercises his powers mostly to better himself. Petitionary prayers clutter his thoughts and annoy him, so he simply answers Yes to all of them. At one point he has a "falling out" with his girl friend, Grace Connelly (played by Jennifer Aniston). Bruce tries to exercise his powers to make her to choose to forgive him. His attempts fail because he cannot ensure that Grace freely chooses to forgive him.
    There is nothing surprising about the moment depicting Bruce's inability to force Grace to forgive him. The film's viewers need not (again) tap into their willing-suspension-of-disbelief. The reason is simple. Forgiveness, in contrast with the supposed forgiveness that consists in merely uttering sounds in response to "I'm sorry," requires that it is up to the

offended individual whether to accept or to reject the apology. Acts of forgiveness intuitively imply freedom.

For the unconvinced, consider the situation from another angle. Suppose that the film had shown Bruce succeed in causing Grace to choose to forgive him. How would the audience respond? The viewers would be surprised and film critics, uncharitable. However, tap into your willing-suspension-of-disbelief by supposing that Bruce succeeds in causing Grace to choose to forgive and love him. Suppose the film continues. They are back together again and Bruce lets Grace in on his little secret (say, after he loses his powers). He tells her that he caused her to choose to forgive him. What response would viewers predict that Grace might make? Might she say, "Well, whence came my choice doesn't matter. I got to do what I chose to do. After all, I did forgive you. I'm delighted you made the effort to plot the course of our lives in this manner"? By no means.[37] Intuitively, forcing another's choice to forgive just doesn't square with our conception of what it means to forgive.

Thus, acts worthy of the name 'forgiveness' are intuitively free. Perhaps the same points apply, *mutatis mutandis*, for certain sorts of love—not the sort of love obtaining only when someone 'falls in love' but rather the kind of deep love that the Greeks called *agape*. Moreover, we cherish forgiveness and love. Poets have taught us this much. Therefore, since forgiveness and love matter, there is a kind of freedom that matters. If this kind of freedom required by forgiveness and love is the same kind of dignifying freedom required by deliberation, the attainability of life-hopes, and moral responsibility, then let this kind be the object of our study.

2.3.6   Reactive Attitudes
Robert Kane (2002a: 5) writes,

> Gratitude, resentment, admiration, indignation, and other such reactive attitudes seem to depend upon the assumption that the acts for which we feel grateful, resentful, or admiring originated in the persons to whom we direct these attitudes. We believe that it was up to them whether they performed those acts or not.

If Kane is right, then free will matters since these reactive attitudes are important to us. What reasons are there for thinking Kane is right?

Consider the admiration we have for certain inventors. Examples of inventions come readily to mind. People have produced artifacts such as Velcro, the pogo stick, the hinge, the bicycle, the flushing toilet, the match, the light bulb, the combustion engine, the shoe, and the space shuttle. In typical cases of invention, the inventor produces an artifact that serves a purpose. An artifact's purpose or function implies that there is a design. The most likely suspect for being the source of genuine (as opposed to merely apparent) design is an agent who possesses intentions about how the artifact is supposed to work.

Sometimes an artifact's design is merely apparent, but in these cases we do not attribute the design to an agent. For example, Ivory soap floats. Is it supposed to float? Yes. However, the fact that soaps can float was discovered by accident. As things happened to have turned out, a factory worker, responsible for regulating huge vats churning a liquid concoction prior to

---

[37] And even if we express the same propositions without using the term 'whence' (after all, it is a popular film), the answer remains the same: by no means.

solidifying it into soap bars, neglected his duties in order to get a bite to eat for lunch. Consequently, the liquid mixture got overly aerated—puffed-up and frothy. To everyone's surprise, the soap bars floated. The original, uncared-for batch was not supposed to float. The negligent factory worker is not really responsible for the buoyant disposition of Ivory soap, even though perhaps he could have refrained from neglecting his duties (by stopping the mixing machine during his lunch break, say). However, someone was smart enough to continue to implement the same procedure in order to make floatable soap. Though the first batch of aerated soap did not exactly turn out how *it* was designed, someone refrained from seeing it only as a botched batch. Now, every batch carries the design adored by many river bathers.

The derelict attendant did not in any metaphysically robust sense originate the idea of, or the intention to make, floating soap. He was a passive though fortunate player in the discovery process. By contrast, typical cases of design are cases of invented artifacts, where an agent acts in such as way as to be the source of the artifact along with its intended function.

Being the originator of a designed artifact and a free action's implying origination is no coincidence. In characteristic cases of invention, agents both have the power to invent or not to invent and are originators of their invention. Thus, standard cases of invention jibe neatly with the common belief that many inventors freely come up with their inventions. Were there no free acts of invention, it would be difficult to ascribe an invention to anyone in a way more worthy of ascription than one could ascribe to the negligent factory attendant.[38] Thus, deserved admiration (along with analogous reactive attitudes) seems to require our target kind of freedom.

2.4   Conclusion

If the kind of freedom required by reactive attitudes (such as admiration) is the same kind of dignifying, authorship-inducing freedom required by certain creative acts, forgiveness and love, deliberation, the attainability of life-hopes, and responsibility, then let this kind be the object of our study. If there is such a kind of freedom, it appears tremendously valuable.

Demonstrating that all of these valuable features of our lives hang together and relate, in the ways suggested above, to a single kind of freedom would be extremely difficult. While being careful to avoid the informal fallacy of appealing to authority, it is worth noting that many philosophers who have specialized for decades in the topic of free will believe that there is this strict and demanding sort of freedom.[39] Moreover, it does not seem obviously incoherent. Even if only some of these valuable features were so related to a single kind of freedom, there would be ample reason to suppose that freedom—i.e., that dual, originating power exercised by intelligent agents who act—is valuable. Since I doubt that this conception of freedom can be shown incoherent with a valid argument whose premises are less controversial than the denial of the intuitive beliefs heretofore canvassed from common folk and the wise, I assume this kind of freedom in the rest of this book.

---

[38] For more on the relationships between reactive attitudes and free will, see Peter Strawson (1962).

[39] E.g., Clarke (2003), Kane (1995), and O'Connor (2000a, 1994).

CHAPTER 3

THE COMPATIBILITY QUESTION

*Nor is he free who is what he wants to be, since what a person
wants to be is obviously determined by factors outside him.*
—Rabbi Abraham Joshua Heschel (1955: 409)

This chapter continues to set the stage for subsequent chapters, motivating taking seriously the Intelligibility Question. In the first section, I discuss in detail the type of determinism that most compatibilists and incompatibilists commonly assume in their disputes. The next section, §3.2, makes precise the claim that constitutes the locus of disagreement between compatibilists and incompatibilists. §3.3 proposes that between compatibilism and incompatibilism, the latter is the default, folk belief. This section also roughly outlines the most popular reasons that compatibilists have for giving up their default belief in incompatibilism— reasons that I attempt to undermine in the remainder of my book. Finally, §3.4 investigates the adequacy of Ted Warfield's recent and novel argument for incompatibilism. I defend Warfield's argument from two objections proposed by Dana Nelkin and Samuel Rickless (2002). I then discuss two criticisms that Thomas Flint (1998b) advanced against a pre-published version of Warfield's argument. I explain how Flint's objections raise considerable skepticism against Warfield's project.

The upshot is that even though common people pre-reflectively hold to incompatibilism, believing that there are undetermined free actions. Arguments against compatibilism, though, remain exceedingly controversial. In the following chapters, I turn to assessing the intelligibility of the layperson's conviction. If an undetermined free action is possible, then there should be a fairly rigorous way to make sense of it. To confirm the non-vacuous truth of this proposition, Chapter 4 attempts to undercut Peter van Inwagen's (2001, 2000) recent argument against the possibility of an undetermined free action. The remaining four chapters zero in and defend a fairly rigorous account of an undetermined free action. We'll also see that of the major candidates putatively accounting for an undetermined free action, only agent-causation underwrites in a principled way the layperson's default belief in incompatibilism.

3.1   Determinism
       Up until this point I have assumed that determinism is the thesis that the past determines
a unique future.[1] But what exactly does this mean? Jordan Sobel (1998: 77) ruminates,

> Is determinism compatible with free will? … The answer to it, I maintain, can be
> only, Yes and No. Everything depends upon what one means by 'determinism,'

---

[1] I first mentioned the notion of determinism in §1.2.

and by 'free will.' These are terms of art, and within limits set by somewhat unsettled traditions, we make of them what we want.

In Chapter 2, I discussed at length a popular and intuitive kind of freedom. This section attempts to get clear about what kind of determinism is most relevant to our study, explicating and advocating a commonly accepted account of determinism.

3.1.1   Causal Determinism

For better or for worse, there are many types of determinism. Sobel (1998: 77-166) maintains that there are at least ninety varieties. However, there is somewhat of a recent consensus about how to understand at least one kind of determinism that is relevant to the metaphysical problem of free will. Most analytic philosophers defer to van Inwagen (1983: 3), who characterizes determinism as "…the thesis that there is at any instant exactly one physically possible future."[2] Van Inwagen (1983: 59, 65) further elucidates this definition, letting the term 'determinism' denote the conjunction of the following two theses:

- "For every instant of time, there is a proposition that expresses the state of the [entire physical] world at that instant;

- if p and q are any propositions that express the state of the [entire physical] world at some instants, then the conjunction of p with the laws of nature entails q."[3]

What are the precise conditions for a proposition's expressing the state of the world at an instant? Van Inwagen does not say, taking the notion as primitive. I assume this is a virtue, as I think that niggling over the precise conditions is largely irrelevant to answering the Compatibility Question. Nonetheless, there are two non-negotiable restrictions. Van Inwagen (1983: 59-60) states:

> Provided the following two conditions are met, one may flesh out 'the state of the world' in any way one likes. (i) Our concept of *state* must be such that, given that the world is in a certain state at a certain instant, nothing *follows* about its state at any other instant… (ii) If there is some observable change in the way things are—if a white cloth becomes blue, a warm liquid cold, or if a man raises his hand—this change must entail some change in the state of the world.

Condition (ii) of this quotation seems straightforward. However, grasping the importance of (i) may require more reflection. Presumably, proposition p expresses the state of the world only if p is contingent. So (i) reads as follows: If p and q are any propositions that express the state of the

---

[2] Van Inwagen in turn cites Sidney Hook, ed. (1958). In order to ensure that the kind of determinism in question is future-directed, say that proposition p expresses a state at an instant *prior to* the instant of the state expressed by proposition q.

[3] See also his (1975, 1974). Hereafter, let 'state of the world' abbreviate 'state of the [entire physical] world'.

world at distinct instants, then p does not entail q. That is, if p and q are any propositions that express the state of the world at distinct instants T1 and T2, then there is a possible world including both p's expressing the state of the world at T1 and q's failing to express the state of the world at T2. So, p alone does not entail q. In order to guarantee q—to determine q—, something else is needed in addition to p. If a proposition expressing the laws of nature L is such that (p & L) entails q, then we have a case where determinism holds. The key, here, is that condition (i) makes room for the second conjunct of van Inwagen's definition of determinism to do the work it is supposed to do.

Regarding the phrase 'law of nature', van Inwagen (1983: 60) says, "I have no idea how to explain this term, much less define it." Nonetheless, he thinks 'law of nature' should not be understood as fundamentally epistemological. The truth of a proposition expressing the laws of nature does not depend on what any (contingent) entity thinks. Moreover, van Inwagen (1983: 6) defers to philosophical precedent, suggesting that a law of nature "…is supposed to be true,[4] to be contingent, to entail the existence of no particular (contingent) individual and to 'support its counter-factuals' or 'warrant inference to subjunctive conditionals'."[5] Most important, van Inwagen (1983: 62) stipulates that "…if human beings *can* (i.e., have it within their power to) conduct an experiment or construct a device that would falsify a certain proposition, then that proposition is not a law of nature." Finally, a law of nature depends neither on what any contingent agent does nor on what any contingent agent fails to do.

Is there anything more that we can say about how to understand the notion of a natural law? Presumably, natural laws specify a relation between holding between two events (or types of events) that is general in the right sort of way. But in virtue of what does this relation hold? The answer is not trivial.

Here are some common examples of natural laws: any specimen of gold (Au) melts at 1,063° C, heated metals expand, H_2O boils at 100° C at standard pressure, physical objects attract each other proportionally to the inverse of their distances squared, friction yields heat, copper conducts electricity, and the degree of order in a closed system decreases over time (2nd Law of Thermodynamics). These are general truths of a certain kind. But what kind? The answer is difficult to come by, and the leading candidates are several. I explain some of the major contenders, leaving it open for now which one is best. I believe that any of these approaches to understanding natural laws is consistent with my arguments in this chapter.

---

[4] Elsewhere, van Inwagen never asserts that a proposition expressing a law of nature must be true. Indeed, he (1995b: 48) says, "Some philosophers insist that, by definition, a law of nature, whatever, else it may be, must be a true proposition. I can't think why." Since there does not seem to be any good reason to think it could be that a law of nature is a false proposition, I just *assume* with van Inwagen (1983, 1995b: 47) that the laws of nature in world W are true in W.

[5] That is, from it is a law of nature that all Fs are Gs, we may infer that everything is such that were it to have property F, it would have property G. Van Inwagen (1983: 6) rightly notes that one may not validly infer a law of nature from its corresponding counterfactual. Interestingly, Chisholm (1981a: 61) provides the following sufficient condition for a state of affairs being a law of nature in terms of a counterfactual: If it is true that, for every x, and for every property H such that everything has H, if x were to have both F and H then x would have G, then: it is a law of nature that whatever has both F and H has G.

One type of view, the *Humean* approach, has it that natural laws are just certain regularities.[6] For example, perhaps the right sort of generality inherent in natural laws consists in truth-functional regularities.[7] On this view, it is a law of nature that *x*'s being a specimen of gold heated to 1,063° C materially implies *x*'s melting—that is, either it is not the case that a specimen of gold is heated to 1,063° C, or else the specimen melts.

But this conception of natural laws can't be quite right, since many truth-functional generalizations should not count as laws. For example, it may happen to be true that all the coins in Mr. Goodman's pocket are silver. Coins' being in Mr. Goodman's pocket is constantly conjoined with the coins' being silver. Thus, since *x*'s being a coin in Mr. Goodman's pocket materially implies *x*'s being silver, it is a law of nature that all coins in Mr. Goodman's pocket are silver, which is absurd. Intuitively, a natural law warrants an inference to its corresponding subjunctive conditional, and subjunctive conditionals are not truth-functional.[8] For example, it is not the case that for any coin such that it were Mr. Goodman's pocket, it would be silver. For, if there
were a penny in Mr. Goodman's pocket, it would not be silver. That all the coins in Mr. Goodman's pocket are silver, though in fact a true generalization, is just an accidental truth. Presumably, natural laws are more 'binding' and less 'accidental'.

There are more sophisticated *Humean* approaches. David Papineau (1995: 475) explains:

> Perhaps a better Humean solution is that proposed by F.P. Ramsey, and later revived by David Lewis: laws are those true generalizations that can be fitted into an ideal systematization of knowledge—or, as Ramsey put it, laws are a 'consequence of those propositions which we should take as axioms if we knew everything and organized it as simply as possible in a deductive system'. Accidents are then those true generalizations which cannot be explained within such an ideal theory.

David Lewis (1999: 222) claims that an ideal deductive system

> must be as simple in axiomatization as it can be without sacrificing too much information content, and it must have as much information content as it can have without sacrificing too much simplicity. A law is any regularity that earns inclusion in the ideal system.[9]

Informatively characterizing the notions of degree of simplicity, degree of information content, and how one can maximize each in a deductive system is not trivial.

In contrast to Humean approaches, one might think that the right sort of generality inherent in natural laws consists in their being metaphysically necessary. On this reading, law

---

[6] For another short overview of the Humean approach, see O'Connor (2000a: 68).

[7] See Frederick Suppe (1977) and Harold Brown (1977).

[8] See Lewis (1973).

[9] Fred Dretske (1977) and David Armstrong (1983) expound some difficulties for regularity theories.

statements are categorical necessary truths, which *do* warrant inferences to their corresponding subjunctive conditionals. Surely it is metaphysically possible, however, that gold melts at 1,075° C, or that H_2O boils at 150° C, or that physical objects attract each other proportionally to the inverse of their distances to the 1.99999999 power, or that the speed of light were half of what it is in our world, or that the degree of order in a closed system remains the same over time. Natural laws surely do not have the same status as mathematical truths.[10] For, even though a proposition *p* expressing some natural law holds in some world W, there is a world where *p* does not hold. Thus, natural laws are not necessary truths.

Fred Dretske, Michael Tooley, and David Armstrong (DTA) independently constructed another type of anti-Humean view of natural laws.[11] Yuri Balashov (2002: 460-1, <u>emphasis mine</u>), extracting the shared features of this realist view, describes the view as follows:

> According to this theory, <u>laws of nature are objective relations, not between objects, but between properties or universals they instantiate</u>. What makes a given sample of water have density 1 g/cm^3 is not the fact that all samples of water have this density. Although this general fact trivially implies all its instances, it cannot explain or ground any of them. What does this job is a direct relation of "nomic necessitation" between properties *being water* and *having density 1 g/cm^3*. The facts about the world expressed by laws are not general facts covered by universal regularities of the form '$(x)(Px \supset Qx)$' but atomic facts expressed by singular statements of the sort '$N(P,Q)$', where *N* is the postulated relation of nomic necessitation between universals *P-hood* and *Q-hood*.

One might complain that the DTA approach merely labels the problem by postulating a relationship of lawlike necessitation.[12]

The *Causal Powers Theory* represents the final view—or family of views—of natural law, stating that the source of nomic modality resides not in second-order relations between universals but rather in the essential first-order properties of physical objects.[13] Balashov concisely explains the theory's basics:

> The central claim of the theory is that nomic properties include dispositional, as well as categorical varieties. Dispositional properties and their species—propensities, capacities, tendencies, liabilities, trends—often go by the name of *causal powers*… The possession of such powers disposes their bearers to behave in specific ways or to exemplify other characteristic properties. <u>The laws of nature are none other than the facts about dispositional nomic properties essentially possessed by natural kinds of objects</u>. Given the actually existing

---

[10] See, though, Alfred Freddoso (1986).

[11] Dretske (1977), Tooley (1977, 1987), and Armstrong (1978, 1983, 1997).

[12] See Bas van Fraassen (1989: Chapter 5) for criticisms of the DTA approach.

[13] Advocates of the Causal Powers Theory include John Bigelow, Brian Ellis, and Caroline Lierse (1992), Ellis and Lierse (1994), Ellis (2001), Rom Harré and E.H. Madden (1975), Sidney Shoemaker (1998, 1980), Chris Swoyer (1982), and Evan Fales (1990). Balashov (2002) argues that evidence from particle physics and cosmology clearly favors the Causal Powers Theory over the DTA theory.

natural kinds and their essential nomic features, the laws of nature follow necessarily.[14]

Van Inwagen seems to favor a version of the Causal Powers Theory, for he characterizes a law of nature in terms of the causal powers of the tiniest of physical objects. Van Inwagen (1995b: 47) claims, "A proposition is a law of nature in a possible world *w* if it is a contingent proposition that is true in all possible worlds in which elementary particles *always* have the causal powers they *always or almost always* have in *w*," where an elementary particle has a certain causal power if and only if it possesses "a certain intrinsic capacity to affect the motions of other particles" (1995b: 43).

Many philosophers suggest that van Inwagen's determinism deserves to be called *causal determinism*. Here are just three examples.[15] First, O'Connor (2000a: 3) stipulates that "…causal determinism… is the thesis that there are comprehensive natural laws that entail that there is but *one* possible path for the world's evolution through time consistent with its total state (characterized by an appropriate set of variables) at any arbitrary time." Second, Fischer (1986: 33) states, "Causal determinism is the thesis that, for any given time, a complete statement of the hard facts about the world at that time, together with a complete statement of the laws of nature, entails every truth as to what happens after that time."[16] Third, Michael Zimmerman (1988: 217) says that determinism is, "…roughly, the doctrine that there is a sufficient causal condition of every event,"[17] where C is a sufficient causal condition of E iff C occurs, C does not imply E, and it is physically necessary (i.e., a law of nature) that if C occurs at a time T, then E occurs at or later than T. I believe that Zimmerman and van Inwagen's definitions are equivalent.[18] For, let

---

[14] Balashov (2002: 461, <u>emphasis mine</u>).

[15] For more examples, see Sobel (1998: 98), John Earman (1986: 13), and Randolph Clarke (2003: 4). Though many thinkers speak of *causal* determinism, van Inwagen himself prefers to avoid employing causal notions. He (1983: 65) writes, "The reader will not that the horrible little word 'cause' does not appear in this definition [of determinism]. Causation is a morass in which I for one refuse to set foot. Or not unless I am pushed."

[16] For an enlightening and detailed treatment of the distinction between hard-facts and soft-facts, see William Lane Craig (1991: Chapter 9). He basically takes the position that a true proposition is a hard fact *iff* it would be true even if the future were annihilated. Craig (1991: 189) writes:

[A] fact is soft iff it is a past or present event or actuality which is counterfactually dependent upon some future event or actuality in such a way that the earlier event or actuality is a consequence of which the later event or actuality is the condition. A fact is hard iff it is a past or present event or actuality which is not so dependent. … To say a past fact is soft entails the claim that were some future condition to be other than it will be, then the past fact would as a consequence have been different than it was. By contrast, to say a past fact is hard is to say that no matter how future conditions might vary, the past would have been the same in every case.

[17] Zimmerman follows Chisholm (1976a: 58) in defining a sufficient causal condition, which I discuss in detail in Chapter 6.

[18] Van Inwagen (1986a: 242) recognizes that

…determinism is the thesis that the past and the laws of nature together determine a unique future and is *not* the thesis that every event has a cause ('universal causation'). For the thesis of universal causation might be true and determinism false.

'E_p' denote the occurring event *its coming to pass that* p *is true*, for any true proposition *p* expressing any state of the world. It follows that for any proposition *q* expressing the state of the world at a later moment than the moment *p* is true, (*p* & *L*) entails *q* if and only if (*Ep* occurs and it is a law of nature that if *Ep* occurs then *Eq* occurs) entails *Eq* occurs.

In fine, determinism (D) is "the thesis that there is at any instant exactly one physically possible future."[19] In other words, let *determinism* be the thesis that there is a sufficient causal condition for every event. Say that an event E is *determined iff* there is a sufficient causal condition for E. Say that an event E is *undetermined iff* E is not determined. It is very important to note that an undetermined event could still be caused. For example, it could be that some event C indeterministically or probabilistically causes event E.[20]

3.1.2   The Relevancy of van Inwagen's Determinism

Van Inwagen's determinism is rigorous and, upon reflection, straightforward. Moreover, showing that this kind of determinism precludes anyone from freely acting would be a serious accomplishment. One may argue, though, that van Inwagen's employing this kind of determinism to argue that free will is incompatible with determinism is like using an atomic bomb to eliminate a dust mite. Stated less metaphorically, van Inwagen's rendering of determinism is far stronger than it needs to be. For, there are intuitive cases where one's performance of an action is determined even though van Inwagen's determinism is false. So, even if free will is incompatible with van Inwagen's determinism, might there be conceptual space leaving room for the compatibilist to claim that free will may be compatible with a weaker, perhaps more relevant, variety of determinism? Translating this question into an objection and making it stick is not trivial. I argue in this section that two ways of filling in this objection fail to stick.

First, suppose that from the moment right before Sam Slug is born there is a sufficient causal condition for every action Sam ever performs. However, suppose that one such sufficient causal condition does not itself have a sufficient causal condition. Perhaps there is some random event that antedates but causally contributes to Sam's coming into existence. Or perhaps the random event occurs shortly after the Big Bang. Suppose, e.g., that were the random event in the immediate wake of the Big Bang not to occur, Sam would not exist. Even so, intuitively, every action Sam performs is determined, since each action has a sufficient causal condition. However, it is not the case that if *p* and *q* are any propositions that express the state of the entire physical world at some instants, then the conjunction of *p* with the laws of nature entails *q*. Thus, intuitively, all of one's actions may be intuitively determined even though van Inwagen's determinism is false.

Notice that one can locate a random event in many different places and still make the same point. Here are a few examples. Imagine that Sam performs action A. Suppose that even

---

For, the cause *C* of event *E* might not be a sufficient causal condition of *E*. Van Inwagen (1983: 129, 139-141) cites G.E.M. Anscombe (1998). For van Inwagen's argument that universal causation does not imply determinism, see van Inwagen (1983: 3-5, 1977: 89, 90).

[19] Van Inwagen (1983: 3).

[20] For more in support of this critical distinction, see Anscombe (1998), van Inwagen (1983), Kane (1996a), Clarke (2003), O'Connor (2000a), and Mele (1995).

though there is a sufficient causal condition for Sam's performing A, there occurs at the very same moment that Sam acts a few random events occurring in some nebulae in a very distant galaxy. Intuitively, these particular undetermined events have no bearing on whether or not Sam's action is determined. Or, suppose that there is a sufficient causal condition for Sam's performing A even though a random event occurs within Sam's brain, but where the random event plays no causal role in—does not contribute causally to—Sam's performing A. In each case it seems obvious that what Sam does is determined even though van Inwagen's determinism is false. Thus, says the critic, the kind of determinism relevant to the problem of free will need not be van Inwagen's kind. The critic's point trades on the intuitive belief that if every action has a previously occurring sufficient causal condition, then every action is determined, even though some other event in the physical cosmos has no sufficient causal condition. The performance of an action can be locally determined without occurring in a global deterministic system.

Moreover, having no sufficient causal condition is not the only intuitive way for one's performance of an action to be determined. Suppose, e.g., that Sam's performing A has no sufficient causal condition. Let 'H' denote the hard past, let 'L' denote the laws of nature, and let 'B' denote the entire event *Sam's performing A*. Thus, even though (H & L & B), it is not the case that (H & L) entails B. Hence, ◊(H & L & $\sim B$). Now, suppose that God makes it the case that Sam performs A not by causing a sufficient causal condition of B but by directly causing B to occur. Intuitively, even though B has no sufficient causal condition, B is determined. Intuitively, Sam does not act freely since God forces him perform the action.[21]

I think that the right responses on behalf of van Inwagen are these. With respect to the first objection, it is just obvious that, necessarily, if free will is incompatible with van Inwagen's determinism, then free will is incompatible with sufficient causal conditions of the sort illustrated in the examples above. Equivalently, it is just obvious that, necessarily, if free will is compatible with every action's having a sufficient causal condition, then free will is compatible with van Inwagen's determinism. With respect to whether or not an action is free, there is no relevant difference between an action's having a sufficient causal condition and that action's occurring in a world where van Inwagen's determinism holds. Since, for any possible world W where some action A is determined *qua* having a sufficient causal condition, there is another possible world W* such that (i) W* includes the same sequence of locally determined events and (ii) van

---

[21] This example may be further fleshed out in a number of ways. For example, we could consistently stipulate that B is wholly out of character for Sam. Perhaps it is true that were God not to cause B, then B would not occur. Moreover, the example could again be developed along different lines. For example, B may be characteristic of what Sam does. For all we know, Sam's reasons may contribute causally to B even though God causes B too. Perhaps God contributes causally to B in such a way that Sam's reasons also contribute causally to B. This supposition is not unreasonable, for God does not contribute causally to Sam's performing A in an event-vacuum. E.g., even if God causes B, Sam's birth contributes causally to B as well, else he wouldn't be there as a subject of God's contributing activity. If this is coherent, there is a minimal notion of Sam's exercising dual power in the sense that his own reasons contribute causally to his acting and there is a possible world in which Sam, in the same *antecedent* circumstances, does not perform the action. For, clearly God is free, in those same antecedent circumstances, not to contribute causally to Sam's performing the action in question. However, God's determining activity precludes Sam from originating his action in the relevant way. Intuitively, this is why Sam is not free.

Inwagen's determinism holds in W*.[22] If action *A* is not free in W*, then *A* is not free in W. Thus, even though van Inwagen's determinism is strong, it is not irrelevantly strong. Thus, showing that free will is incompatible with van Inwagen's determinism suffices for showing that free will is incompatible with the relevant action's having a sufficient causal condition. If the critic still balks, then she should provide a principled account distinguishing between being locally determined and occurring in a global deterministic system, where the account sheds light on how a locally determined action could be free while a nearby global deterministic system precludes the corresponding action from being free.

The second objection also fails to stick. Earning the claim that van Inwagen's determinism is insufficiently relevant to the problem of free will requires showing that even if free will is incompatible with van Inwagen's determinism, there is a distinct and intuitive form of determinism that is compatible with free will. However, very few compatibilists would think that Sam could freely perform an action whose performance is directly made to occur by another agent.[23]

3.2   Homing in on the Locus of Dispute
        This section isolates the heart of the incompatibilist position. I introduce some terminology that simplifies expressing the critical point of disagreement between compatibilists and incompatibilists, and I use this terminology in the remainder of my book. This critical point of disagreement also captures precisely *where* the incompatibilist thinks determinism needs to break down in order to leave room for freedom.

        Interestingly, the case on which the last objection in §3.1 rests illustrates a point that incompatibilists take quite seriously. If freedom requires the lack of determinism, then freedom requires more than just a denial of van Inwagen's determinism. According to incompatibilists, necessarily, if freedom requires indeterminism, then an agent's freely performing an action neither has a sufficient causal condition nor is made to occur by some other agent. More precisely, incompatibilists contend that freedom precludes a determined action defined as follows

(DA)   An agent Δ's action A is *determined* =df. Δ performs A, there is no sufficient causal condition for *Δ's performing A* only if there is another agent Δ* that makes Δ perform A.

        The central claim over which incompatibilists and compatibilists disagree is easily understood with a bit of technical terminology. Consider an agent's very first putatively free action. Typically, right before the agent acts, she is in a certain state of mind and has a certain character. She has at this time a specific psychological constellation of beliefs, desires, and values. Many factors causally influencing her first (putatively free) action are at work. *We* could not begin to specify these factors completely. But they constitute a set of detailed circumstances that are nonetheless present and, in some sense hard to specify, relevant to the etiology of which

---

[22] Of course, the laws of each world may be different, but that is not clearly relevant.

[23] I think van Inwagen concurs. Elsewhere he (1999: 54) says, "I take it to be obvious that if God decrees (I do not mean *commands*) that a certain human being on a certain occasion behave in a certain way, then that human being loses his freedom of choice on that particular occasion."

action is performed. A comprehensive specification of these forces would be quite rich, including the entire history of causal factors leading up to the moment right before the time of the action. Call the totality of these factors *antecedently complete circumstances*. Suppose also that simultaneous causal influence is possible. Consider
only the circumstances including all of the causal factors other than the agent that affect the agent *at the same moment* she acts. Call these *complete simultaneous circumstances*. Complete simultaneous circumstances, then, are equivalent to a proposition expressing the causal activity of all other agents at the time the agent acts. Call the combination of complete simultaneous circumstances together with antecedently complete circumstances simply *complete circumstances*. To be safe, complete circumstances include the entire causal history of the world prior to the time of the action, together with whatever outside influences are affecting the agent at the time of her action.

Incompatibilists hold that, necessarily, the complete circumstances in which an agent freely acts do not guarantee precisely which action the agent freely performs. That is, necessarily, if the agent freely performs action *A*, the set of complete circumstances does not include the agent's freely performing *A*. If the set of complete circumstances does not include which action the agent freely performs, say that they are *non-determining* circumstances. Thus, incompatibilists maintain that an agent acts freely only in complete *nondetermining* circumstances.[24] Incompatibilists may hold that an agent acts in complete determining circumstances, but since they are determining, the act is not free. I take it for granted that an unfree action is possible.[25]

Compatibilists, on the other hand, contend that it is possible for an agent to act freely in complete determining circumstances. Thus, incompatibilists believe and compatibilists reject the following proposition:

(Crux)   Necessarily, if an agent Δ freely performs some action A, then the set of complete circumstances in which Δ acts does not include Δ*'s freely performing A*.

3.3   Default Beliefs and Hors d'oeuvres
This section proposes that ordinary folk have a default belief in incompatibilism. In addition, since giving up some of one's default beliefs is sometimes reasonable and appropriate, I outline some of the most popular reasons that a compatibilist may give for why one should abandon one's default belief in incompatibilism. In the remainder of my book, I attempt to undermine these reasons.

I am inclined to think that most people pre-reflectively find incompatibilism more credible than compatibilism. Over the past several years I have had many exchanges with well-educated yet philosophically untutored people. When asked whether one's very first free action at a moment could be guaranteed to occur by previously occurring events, most people pre-reflectively answer No. We believe that at the moment one freely performs an action *A* in a set C of circumstances one could just as well freely refrain from doing *A* in the same set of

---

[24] I've borrowed from Thomas Flint (1998a: 47, 200) this characterization of what counts as complete circumstances that are nondetermining with respect to which action an agent performs.

[25] Contrary to Richard Taylor (1992).

circumstances. And we (rightly or wrongly) pre-reflectively believe that were one to refrain freely in C, then it would have been in one's power to do *A* freely in *C*. Most people have no argument for incompatibilism, yet they are incompatibilists. Their belief in incompatibilism is basic. Van Inwagen (1993: 187) confirms this observation:

> It has seemed obvious to most people who have not been exposed (perhaps 'subjected' would be a better word) to philosophy that free will and determinism are incompatible. It is almost impossible to get beginning students of philosophy to take seriously the idea that there could be such a thing as free will in a deterministic universe. Indeed, people who have not been exposed to philosophy usually understand the word 'determinism' (if they know the word at all) to stand for the thesis that there is no free will. And you might think that the incompatibility of free will and determinism deserves to seem obvious—because it is obvious.

If my many conversations with lay people and van Inwagen's experience as an undergraduate instructor fairly indicate the status of folk opinion on the Compatibility Question, then it seems that one can rationally assume the reasonability of incompatibilism until this assumption is shown mistaken. Thomas Reid agrees that assuming incompatibilism from the outset is dialectically appropriate. Reid (1895: 313-314) states:

> This natural conviction of our acting freely, which is acknowledged by many [i.e., compatibilists] who hold the doctrine of necessity, ought to throw the whole burden of proof upon that side; for, by this, the side of liberty [i.e., incompatibilism] has what lawyers call a *jus quaesitum*, or a right of ancient possession, which ought to stand good till it be overturned.

Now one may object, "Why not just let your assumption set—outlined in §1.6—include the thesis of incompatibilism? After all, if it is so obvious, then just assume it and move on to address the Intelligibility Question?"

The reason I have not simply assumed incompatibilism is fairly simple. The truth of incompatibilism is too controversial and yet bears directly on the topic of free will. Reflecting on our pre-reflective beliefs often reveals the controversial nature of those beliefs. Reasonable people sometimes change their minds, rejecting previous beliefs upon entertaining new considerations. In connection with our topic, many people have changed their minds regarding the Compatibility Question, which underscores the contention that pre-reflective folk belief is not inflexibly authoritative.

So, under what conditions is it dialectically appropriate to abandon one's intuitive, pre-reflective belief in incompatibilism? Let's put the question more generally. Under what conditions is it dialectically appropriate to abandon an arbitrary intuitive, pre-reflective belief? Roughly put, what normally occurs when reasonably abandoning an arbitrary belief?[26] Usually, people rationally give up a pre-reflective belief upon reflecting on new information. When this new information is just as intuitive as (or more intuitive than) the pre-reflective belief, and yet

---

[26] Rigorously answering this question would take my study too far afield. I trust that the gist of what I'm saying comes through.

the new information clearly implies the falsity of the previously held belief, one is reasonable in giving up the pre-reflective belief.[27] (Obviously, continuing to hold what was once a pre-reflective belief after reflection means that the belief ceases bearing the property *being pre-reflective* as it acquires the property *being reflective*.) So, on the basis of what sort of new information does one typically and rationally give up one's belief in incompatibilism? Normally, a person rationally abandons her pre-philosophical belief in incompatibilism when either:

(i)   she reflects carefully on compatibilist accounts of free will and finds them at least as plausible as her initial belief in incompatibilism, or

(ii)   she reflects carefully and rationally comes to believe that she should have positive reasons for incompatibilism and finds the best reasons for incompatibilism clearly inadequate, or

(iii)   while believing that acting freely is at least possible, she reflects carefully on what it would be like for an agent's action *not* to be determined by the complete circumstances in which the agent acts, and then rationally thinks that these situations preclude free will; that is, she gives up belief in incompatibilism after rationally coming to believe that there cannot be an adequate answer to the Intelligibility Question.

I do not claim that these conditions exhaust the conditions sufficient for one's rationally abandoning one's pre-reflective belief in incompatibilism. Neither do I claim that these conditions are necessary. However, this is the stuff of which most compatibilists are made. I conclude this section with a brief note on each of (i)-(iii) above, which delimits and circumscribes the general goals of the rest of my book—a taste of what's to come.

    Consider (i). A thorough treatment of the Compatibility Question would include providing a reasoned evaluation of the best compatibilist accounts of free will. For example, the incompatibilist would do well to provide a good argument for what motivates Professor Anscombe's (1998: 256) assertion, "The [compatibilist] reconciliations have always seemed to me either to be so much gobbledegook, or to make the alleged freedom of action quite unreal." While what seems to Anscombe may be veridical, I shall not discuss in detail any compatibilist account. Consequently, my treatment of the Compatibility Question will be less than thorough. I mention (i) only to ignore it. However, I will address the remaining two points that uphold compatibilist convictions—(ii) and (iii).

    Consider (ii). I have encountered no good reason to think that reasonable incompatibilists should at the outset have a positive argument for incompatibilism. Again, incompatibilism seems to be a default conviction. Of course, incompatibilists would be better off were they to have a good argument for incompatibilism. Many incompatibilists believe that the popular Consequence Argument fits this bill.[28] While I register my agreement with Ted Warfield's (2000: 168) claim that "…compatibilist replies to [various sophisticated

---

[27] Strictly speaking, things are more complicated. We should also consider the subject's various background beliefs that are intuitive. Maybe some of these background intuitive beliefs are inconsistent with the new information. The reader is invited to insert a charitable *ceteris paribus* clause.

[28] See §1.2.

Consequence] arguments have been, on the whole, quite weak…," I shall add nothing interesting to this hackneyed debate here.[29] However, the next section, §3.4, investigates a wholly different argument for incompatibilism advanced by Warfield (2000). I rehearse his argument, defend it from a few objections propounded by Dana Nelkin & Samuel Rickless (2002), but conclude that Thomas Flint (1998b) raises considerable doubt against its success.

Consider (iii). Many people "switch over" to compatibilism after contemplating the various alternatives and believing that each of them is clearly less intuitive than compatibilism. Assuming a free action is possible, some scholars hold that an *un*determined free action is impossible. Chapter 4 criticizes van Inwagen's argument for the claim that an undetermined free action is impossible. Some philosophers find inadequate the theories attempting to make sense of an undetermined free action. Chapter 5 argues that one of these theories, *viz*., agent-causation, has an edge over the others. The remaining chapters attempt to defend agent-causation from several objections. Interestingly, I'll discuss how only agent-causation accounts for one's pre-theoretic commitment to incompatibilism *in a principled way*. I conclude, at least by way of inference to the best explanation, that there is a plausible positive answer to the Intelligibility Question. So, if I am right, there is a viable incompatibilist theory of free will: agent-causation.

3.4   Investigating Ted Warfield's New Argument for Incompatibilism
I have already noted that incompatibilism is the folk's default belief. While I do not think the incompatibilist is obliged to have a good reason for her position, it would only help. So, is there a good argument for incompatibilism? Many philosophers contest that the *Consequence Argument* fits the bill, which, to review, van Inwagen (1983: v, 16, 56, 222) nicely distills as follows.

> If determinism is true, then our actions are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are.   Therefore, the consequences of these things (including our present actions) are not up to us.[30]

Of course, many philosophers criticize the Consequence Argument. And then there are the replies to the criticisms, and further criticisms, and criticisms of the further criticisms.[31] This debate seems to have reached an impasse,[32] and I will add nothing that clearly tips the scales.

---

[29] For some interesting recent contributions, see Crisp & Warfield (2000), Finch & Warfield (1998), Fischer (1986c, 1994b: 23-66), Fischer & Ravizza (1998), Ginet (1966, 1990: Chapter 5), Kane (1996a: Chapter 4), Lamb (1977), McKay & Johnson (1996), Nelkin (2001), O'Connor (2000: 5-18, 1993b), Ravizza (1994), Slote (1982), Stump & Fischer (2000), van Inwagen (1983: Chapter 3, 1975), Warfield (2000, 1999, 1996), Widerker (1987), and Wiggins (1973).

[30] See also M.R. Ayer (1968), C. D. Broad (1952), Roderick Chisholm (1976a, 1958), Austin Farrer (1967), Carl Ginet (1966), Anthony Kenny (1975), J.R. Lucas (1970), A.I. Melden (1961), Richard Taylor (1992), David Wiggins (1973), and van Inwagen (1983, 1975).

[31] For discussion of the Consequence Argument and intimately related notions, see Laura Ekstrom (1998, 1995), Finch & Warfield (1998), Fischer (1994, 1983), Fischer & Ravizza (1998: Chapter 6), Richard Foley (1979), Andre Gallois (1977), Ginet (1990), Greenspan (1976), Terence Horgan (1985),

Nonetheless, there is a refreshingly new attempt to argue for incompatibilism, one advanced by Ted Warfield (2000). Warfield's argument for incompatibilism has two stages. First, he advances a pivotal and quite controversial proposition, *viz.*, necessarily, anything anyone is free to make the case is consistent with the hard past. Warfield argues that this controversial proposition strictly implies incompatibilism. In his second stage, Warfield deploys an argument for this very controversial proposition. In particular, he argues that an obviously valid inference requires the truth of this controversial proposition. The obviously valid inference is this: from *P is true and there's nothing anyone is free to do in the circumstances that even might result in ~P* infer *P is true and there's nothing anyone is free to do in the circumstances that would definitely result in ~P.*[33]

Nelkin & Rickless (2002) object to Warfield's argument at each stage. I argue, though, that neither objection succeeds. I then consider two more criticisms of Warfield's second stage advanced by Thomas Flint (1998b) in an unpublished manuscript. I explain why Flint's critique is most damaging.[34]

Let's turn, now, to Warfield's argument. We first assume the following abbreviations:

- *D*: Determinism
- *H*: the complete state of the world in the distant past *with* the laws of nature[35]
- *Fab*: *a* is free to make it the case that *b*

Tomis Kapitan (1996, 1991a, 1991b, 1986), David Lewis (1981), McKay & Johnson (1996), Jan Narveson (1977), O'Connor (2000a: 9-14, 1993), Christopher Peacocke (1999: Chapter 4), Ravizza (1994), van Inwagen (2000, 1994, 1983), Kadri Vihvelin (1995a, 1995b, 1991), Warfield (2000, 1996), and Widerker (1987).

[32] Others agree with my assessment. See Robert Kane (1996a: 14, 16) and Timothy O'Connor (2000a: 5, 17; 1995a: 4).

[33] Warfield (2003: 627-8) presents the following, less rigorous version of his (2000) argument:
(P1)   Necessarily, if determinism is true and one doesn't do *X*, one's doing *X* is inconsistent with the past and laws.
(P2)   Necessarily, if one is free to do *X* then one's doing *X* is consistent with the past and laws.
(C1)   So, necessarily, if determinism is true, then one is not free to do anything other than what one does.
(P3)   Necessarily, if one is not free to do anything other than what one does, one is not free at all.
(C2)   So, necessarily, if determinism is true, then one is not free at all.
I think it is fairly clear that the compatibilist would deny (P2), which just is Warfield's (2000) controversial proposition mentioned above.

[34] Michael Kremer (2004) also contests Warfield's argument. Alas, I discovered this essay too late to give it a fair hearing here.

[35] Warfield (2000: 178) notes,
By 'the complete state of the world' I mean, of course, only the complete *hard* past of the world (excluding, for example, true future tensed propositions). There is, unfortunately, no settled philosophical account of just what features of a time are the 'hard' features of a time.
I again refer the interested reader to Craig (1991: Chapter 9)—see §3.1.1.

3.4.1   Warfield's First Stage
        Warfield (2000: 172-3) proposes the pivotal and quite controversial proposition

(1)   □ ∀s∀x{Fsx ⊃ ◊(H & x)}

and characterizes *determinism* as "the thesis that the conjunction of the [hard] past and laws implies all truth." Determinism, then, implies the truth of

(2)   □ ∀x{D ⊃ (x ⊃ □(H ⊃ x))}.

        We again note, for the record, that (1) is very controversial. It seems that no seasoned compatibilist would accept it. One may think that (1) itself expresses the thesis of incompatibilism. Be that as it may, there is a clearer formulation of the thesis of incompatibilism, and Warfield contends that the clearer formulation follows from (1) and (2). Warfield's first stage just consists in earning this contention. So, I invite the reader to bracket any reservations she may have with (1) until we evaluate in §3.4.2 Warfield's argument for (1).
        So, the first stage of Warfield's project consists in showing that (1) and (2) obviously entail the thesis of incompatibilism, where he characterizes the latter as

(I*)   □ ∀s∀x{(D & x) ⊃ ~Fs~x}.[36]

        Nelkin & Rickless (N&R) agree that (1) and (2) together entail (I*). However, they (2002: 105) argue:

> The problem with this argument is that (I*) does not capture the thesis of incompatibilism (I). What (I*) says is that, necessarily, if determinism is true then every truth is such that no one is free to make it false. But (I) is the view that determinism and freedom are incompatible, i.e. that, necessarily, if determinism is true then no one is free to make anything true.

So, N&R propose that the proper construal of incompatibilism is not (I*) but rather

(I)   □ ∀s∀x{D ⊃ ~Fsx}.

N&R (2002: 105) assert, "Since (I*) is not the same as (I), Warfield's claim that (I*) captures the thesis of incompatibilism is mistaken." Moreover, N&R reason that a seasoned compatibilist may have reasons to deny the move from (I*) to (I), that is, from

---

[36] This argument is straightforward:

(1)	□ ∀s∀x{Fsx ⊃ ◊(H & x)}	[Assumption: Warfield's premise]
(2)	□ ∀x{D ⊃ (x ⊃ □(H ⊃ x))}	[Assumption: Determinism]
(3)	□ ∀x{(D & x) ⊃ ~◊(H & ~x)}	[Equivalent to (2)]
(19)	□ ∀s∀x{Fs~x ⊃ ◊(H & ~x)}	[Equivalent to (1)]
(20)	~◊∃s∃x{(D & x) ⊃ Fs~x}	[From (3) and (19)]
(I*)	□ ∀s∀x{(D & x) ⊃ ~Fs~x}	[Equivalent to (20)]

(I*)   □ ∀s∀x{(D & x) ⊃ ~Fs~x}

to

(I)   □ ∀s∀x{D ⊃ ~Fsx}.

One might wonder how a compatibilist may reasonably deny this inference. According to N&R it is because the inference from (I*) to (I) requires the truth of

(DP)   □ ∀s∀x{Fsx ⊃ Fs~x},

and Frankfurt-style cases may show that (DP) is false.[37] N&R (2002: 105) assert that some philosophers draw "…the conclusion from Frankfurt's example that even if one is not free to act otherwise than one does, one can nevertheless act freely."[38]

  I think that (DP) is true and defensible, but I will not defend it any more than I already have. (DP) seems to express the condition of dual ability, which freedom intuitively requires. I attempted to disable any Frankfurt-style criticisms in Chapter 2, §2.3.1. Moreover, I believe that many compatibilists would agree with (DP) as well, though they would have a different account of the nature of this ability to do otherwise.

  Nevertheless, merely for sake of argument, I concede to N&R that (I*) may not obviously characterize the thesis of incompatibilism and that the move from (I*) to (I) is sufficiently controversial to weaken Warfield's argument for incompatibilism. More precisely, assume that a compatibilist may reasonably deny that (DP) is obviously true. To be clear, I do not concede the falsity of (DP). Rather, I concede that the truth of (DP) is too controversial to be used as a premise for incompatibilism. However, this concession does not guarantee that (1) and (2) fail to entail incompatibilism.

  Moreover, I have a good argument to think that (1) and (2) entail incompatibilism. In addition to the above abbreviations, my argument makes use of another:

- *Mab: a* freely makes it the case that *b.*

I take incompatibilism to be the thesis that, necessarily, if determinism is true then no one freely performs any action. That is, incompatibilism is the thesis that, necessarily, determinism is true only if no one freely makes anything the case. Thus, incompatibilism is true just in the case that

(I**)   □ ∀s∀x{D ⊃ ~Msx}.

My (I**) seems weaker than N&R's (I). Since □ ∀s∀x{Msx ⊃ Fsx}, it follows that (I) entails (I**) but (I**) does not *obviously* entail (I). However, (I**) seems as good a construal of incompatibilism as (I). Compatibilists who hold that acting freely requires that the action be determined would agree that (I**) captures incompatibilism. Incompatibilists maintain that a

---

[37] (DP) appears to express the condition of dual power.

[38] N&R cite Frankfurt (1969) and Mele (1995: 141).

free act is undetermined, and this is just what (I**) says: someone's freely making something the case entails the falsity of determinism.

Now, I maintain that (1) and (2) obviously entail (I**). Here is my argument:

(1) $\Box\ \forall s\forall x\{Fsx \supset \Diamond(H \,\&\, x)\}$ [Warfield's premise]

(2) $\Box\ \forall x\{D \supset (x \supset \Box(H \supset x))\}$ [Determinism]

(3) $\Box\ \forall x\{(D \,\&\, x) \supset \sim\Diamond(H \,\&\, \sim x)\}$ [Equivalent to (2)]

(4) $\Box\ \forall s\forall x\{(D \,\&\, Msx) \supset \sim\Diamond(H \,\&\, \sim Msx)\}$ [From (3)][39]

(5) $\Box\ \forall s\forall x\{Fs(\sim Msx) \supset \Diamond(H \,\&\, \sim Msx)\}$ [From (1)]

(6) $\Box\ \forall s\forall x\{(D \,\&\, Msx) \supset \sim Fs(\sim Msx)\}$ [From (4), (5)]

(7) $\Box\ \forall s\forall x\{Fsx \supset Fs(\sim Msx)\}$ [Basic premise]

(8) $\Box\ \forall s\forall x\{(D \,\&\, Msx) \supset \sim Fsx\}$ [From (6), (7)]

(9) $\Box\ \forall s\forall x\{Msx \supset Fsx\}$ [Basic premise]

(10) $\Box\ \forall s\forall x\{(D \,\&\, Msx) \supset \sim Msx\}$ [From (8), (9)]

(11) $\Diamond\ \exists s\exists x(Msx)$ [Basic premise]

(I**) $\Box\ \forall s\forall x\{D \supset \sim Msx\}$ [From (10), (11)]

My argument is obviously valid. I claim that (1) and (2) together entail (I**). But since my argument clearly smuggles in three underived (i.e., basic) premises, a critic may fault my argument by finding a counterexample to any one of them. They are the following:

(7) $\Box\ \forall s\forall x\{Fsx \supset Fs(\sim Msx)\}$

(9) $\Box\ \forall s\forall x\{Msx \supset Fsx\}$

(11) $\Diamond\ \exists s\exists x(Msx)$

---

[39] The move from (3) to (4) is straightforward. Suppose (3). That is, assume that $\Box\forall x\{(D \,\&\, x) \supset \sim\Diamond(H \,\&\, \sim x)\}$. Consider an arbitrary world W, where $\forall x\{(D \,\&\, x) \supset \sim\Diamond(H \,\&\, \sim x)\}$. By $\forall$-elimination, letting $x = Mba$, infer $\{(D \,\&\, Mba) \supset \sim\Diamond(H \,\&\, \sim Mba)\}$. Since b and a are arbitrarily chosen, infer $\forall s\forall x\{(D \,\&\, Msx) \supset \sim\Diamond(H \,\&\, \sim Msx)\}$. Thus, in world W, $\forall s\forall x\{(D \,\&\, Msx) \supset \sim\Diamond(H \,\&\, \sim Msx)\}$. Since W is arbitrary, infer (4). That is, infer $\Box\forall s\forall x\{(D \,\&\, Msx) \supset \sim\Diamond(H \,\&\, \sim Msx)\}$. Hence, (3) entails (4).

So, any damaging objection must show the falsity of either (7), (9), or (11). Proposition (9) is a conceptual truth, as *freely making* it the case that so-and-so strictly requires *being free to make* it the case that so-and-so. That is, it is impossible that one freely make it the case that so-and-so while not being free to make it the case that so-and-so. Proposition (11) follows both from the thesis of compatibilism as well as from the possibility of libertarianism. For, since freedom is compatible with (in)determinism, then *a fortioria* freedom is compatible with itself. Thus, $\Diamond\exists s\exists x(Msx)$.

Therefore, the critic is down to only one option: she must challenge the truth of (7). Proposition (7) reads, 'Necessarily, for any agent-proposition pair (s, x) such that s is free to make it the case that x, s is free to make it the case that s does not freely make it the case that x'. (7) is not obviously false. I find (7) quite intuitive, and I have yet to encounter any clear counterexample.

Notice that (7) resembles (DP). However, (7) does not fall prey to Frankfurt-style counterexamples, as they only confirm (7). Why? Recall my conclusions in §2.3.1.4. To review the basic point, even if there is a counterfactual manipulator in the shadows, I may still be free to make it the case that b and free to make it the case that I don't freely make it the case that b. For example, suppose Black merely lurks in the shadows as I freely make it the case that b. Now consider the counterfactual case where Black intervenes. He sees that I will not freely make it the case that b and so directly causes me to make it the case that b. Intuitively, even in this counterfactual situation where I do not freely make it the case that b, I am free to make it the case that I do not freely make it the case that b. In the counterfactual situation, Black does not guarantee that I do not freely make it the case that b. And since Black does not guarantee that I do not freely make it the case that b, he in no way precludes my being free to make it the case that I do not freely make it the case that b. Indeed, Black intervenes precisely because he *sees* that I will not freely make it the case that b.[40]

3.4.2   Warfield's Second Stage

Warfield argues ingeniously that an obviously valid inference requires (1). So, since (1) entails incompatibilism, incompatibilism is true. Where the phrase "in the circumstances" is stipulatively equivalent to "given the [hard] past and laws of nature," the obviously valid inference moves from

(12)   *P* is true and there's nothing anyone is free *to do in the circumstances* that even might result in ~*P*

to

---

[40] I'm inclined to think that $\Box\forall s\forall x\{Fs{\sim}x \supset Fs(\sim Msx)\}$ is true. This may *help* explain why one might find (DP) *prima facie* plausible but then, after considering Frankfurt-style examples, think (DP) is too controversial to be a load-bearing premise in any good argument for incompatibilism. For, it is reasonable to think that, for any s and any y, $Fs{\sim}y$ implies but is not implied by $Fs(\sim Msy)$. So, $\forall s\forall y(Fs{\sim}y)$ is stronger than $\forall s\forall y(Fs(\sim Msy))$. This confirms how it may be that some proposition p (say, $\forall s\forall y(Fsy)$) implies $\forall s\forall y(Fs(\sim Msy))$ but does not imply $\forall s\forall y(Fs{\sim}y)$.

(13)   *P* is true and there's nothing anyone is free *to do in the circumstances* that would definitely result in ~*P*.[41]

Warfield maintains that (12) and (13) should be understood as (14) and (15), respectively:

(14) P & $\forall s \forall x \{Fsx \supset \Box((x \ \& \ H) \supset P)\}$

(15) P & $\sim\exists s \exists x \{Fsx \ \& \ \Box((x \ \& \ H) \supset \sim P)\}$.

Warfield (2000: 174-6) argues that if (1) is false, then there is a possible world where (14) holds but (15) does not.[42] But since (14) and (15) correctly translate (12) and (13), respectively, and since there is no possible world where (12) holds and (13) does not, (1) must be true.[43]

---

[41] Warfield (2000: 178-9 notes 12 and 18, *my emphasis*).

[42] So, Warfield argues that (14) entails (15) only if (1). The strategy is to assume the denial of (1) and infer that there is a possible world where (14) is true but (15) is false. Let's see how Warfield's argument works. To review the raw materials, here are (1), (14), and (15):

    (1) $\Box \ \forall s \forall x \{Fsx \supset \Diamond(H \ \& \ x)\}$
    (14) P & $\forall s \forall x \{Fsx \supset \Box((x \ \& \ H) \supset P)\}$
    (15) P & $\sim\exists s \exists x \{Fsx \ \& \ \Box((x \ \& \ H) \supset \sim P)\}$

The argument's structure is broken down as follows. The strategy in lines (21)-(31) is to show that an instance of (14) is true in a world *w* where (1) is false. The strategy in lines (32)-(38) is to show that the corresponding instance of (15) is false in *w*. Lines (39)-(43) complete the deduction. Now for the argument.

(21)	$\Diamond\exists s\exists x\sim\{Fsx \supset \Diamond(H \ \& \ x)\}$	[Assumption: equivalent to the denial of (1) for conditional proof]
(22)	$\Diamond\exists s\exists x\{Fsx \ \& \ \sim\Diamond(H \ \& \ x)\}$	[Equivalent to (21)]
(23)	$\Diamond\{Fba \ \& \ \sim\Diamond(H \ \& \ a)\}$	[Instantiating (22)]
(24)	In *w*, $\{Fba \ \& \ \sim\Diamond(H \ \& \ a)\}$	[Equivalent to (23), $\Diamond$-elimination]
(25)	In *w*, $\sim\Diamond(H \ \& \ a)$	[From (24)]
(26)	In *w*, $\Box(H \supset \sim a)$	[From 25]
(27)	In *w*, H	[From (24): there is a hard past when *b* is free to make *a*]
(28)	In *w*, $\sim a$	[From (26), (27)]
(29)	In *w*, $\forall x \ \Box\{(x \ \& \ H) \supset \sim a\}$	[Trivially from (26)]
(30)	In *w*, $\sim a$ & $\forall s\forall x\{Fsx \supset \Box((x \ \& \ H) \supset \sim a)\}$	[From (28), (29)]
(31)	In *w*, an instance of (14) is true	[From (30): let $P = \sim a$]
(32)	In *w*, Fba	[From (24)]
(33)	In *w*, $\Box((a \ \& \ H) \supset a)$	[From (25) or from $\Box(a \supset a)$]
(34)	In *w*, Fba & $\Box((a \ \& \ H) \supset a)$	[From (32), (33)]
(35)	In *w*, Fba & $\Box((a \ \& \ H) \supset \sim P)$	[From (34): let $P = \sim a$]
(36)	In *w*, $\exists s\exists x\{Fsx \ \& \ \Box((x \ \& \ H) \supset \sim P)\}$	[From (35): $\exists$-introduction]
(37)	In *w*, $\sim(P \ \& \ \sim\exists s\exists x\{Fsx \ \& \ \Box((x \ \& \ H) \supset \sim P)\})$	[From (36)]
(38)	In *w*, $\sim$(15)	[Equivalent to (37)]
(39)	In *w*, (14) and $\sim$(15)	[From (31), (38)]
(40)	If (21), then $\Diamond\{$(14) & $\sim$(15)$\}$	[Completes conditional proof, (21)-(39)]
(41)	If $\sim$(1), then $\Diamond\{$(14) & $\sim$(15)$\}$	[Equivalent to (40)]
(42)	If $\sim\Diamond\{$(14) & $\sim$(15)$\}$, then (1)	[Equivalent to (41)]

3.4.2.1   Nelkin & Rickless's Objection

      N&R concede that (14) implies (15) only if (1) is true. They also do not dispute that (15) represents (13). However, they think that (14) incorrectly translates (12). They (2002: 106) declare, "The correct translation of [(12)] is [(14*)]:

(14*) $P \ \& \ \sim\exists s\exists x\{Fsx \ \& \ \Diamond((x \ \& \ H) \supset \sim P)\}$."

Notice that (14*) is equivalent to

(14**) $P \ \& \ \forall s\forall x\{Fsx \supset \Box((x \ \& \ H) \ \& \ P)\}$.

      N&R underscore two points. First, they note that (14**) implies but is not implied by Warfield's (14). Hence, even if (14) requires (1) to imply (15), it may be that (14**) does not require (1) to imply (15). Second, they see that (14**) obviously implies (15) without (1). In order to see these two obvious points at a glance, here are the crucial propositions—annotated accordingly.

(14**) $P \ \& \ \forall s\forall x\{Fsx \supset \Box((x \ \& \ H) \ \& \ P)\}$ [Equivalent to (14*), i.e., to N&R's translation of (12)]

(14) $P \ \& \ \forall s\forall x\{Fsx \supset \Box((x \ \& \ H) \supset P)\}$ [Warfield's translation of (12)]

(15*) $P \ \& \ \forall s\forall x\{Fsx \supset \Diamond((x \ \& \ H) \ \& \ P)\}$ [Equivalent to Warfield's (15)]

N&R (2002: 106) conclude,

> That [(14**)] entails [(15*)] follows directly from the fact that a proposition's being necessarily true entails that it is possibly true. This fact, and not the truth of [(1)], is what explains the [valid inference from (12) to (13)]. We conclude that Warfield's argument for the truth of [(1)] fails.

      The crux of the dispute, then, hinges on which of (14) or (14*) translates (12). In what follows I discuss how Warfield has the upper hand on two points, dialectically speaking. I will argue that N&R's translation of (12) fails, concluding that Warfield's argument remains unchallenged.

      Warfield's case seems dialectically superior on at least two points. First, N&R merely stipulate that their (14*) translates (12). It would be nice were they to offer some intuitive support on behalf of understanding (12) as (14*). In contrast, Warfield offers an explanation for his translation. He says,

---

(43)   If $\Box\{(14) \supset (15)\}$, then (1), Q.E.D.   [Equivalent to (42)]

[43] Hereafter, let 'translates' abbreviate 'correctly translates'.

[(12)], translated as [(14)], says that P is true and every agent-proposition pair is such that if the agent is free to bring about the proposition then P is going to (still) be true even if the proposition is 'added' to the past and laws. This is the clearest understanding of a proposition, the '~P' of [(12)], being such that (given the circumstances) no one is free to make it the case.[44]

Perhaps this explanation is meager, but it's nonetheless something. N&R neither offer an explanation of their own translation, nor do they challenge Warfield's explanation for his translation. The degree of plausible support Warfield gives for his translation is, *ceteris paribus*, the degree to which it has a slight edge over N&R's translation.

Second, if N&R are right, then one should be able to offer a clear case where Warfield's (14) holds while (12) does not. N&R provide no grounds whatsoever for thinking it could be the case that (14) is true while (12) is false. The lack of a direct counterexample to Warfield's translation of (12) weakens N&R's case against Warfield's argument for incompatibilism.

Regardless of whether (14) translates (12)—and I see no compelling reason to the contrary—there is a good reason to think N&R's (14*) does not translate (12). Notice that (14*) is quite strong. It entails (12), (14), and (15). However, since (14*) is supposed to translate (12), they should each imply the other. But here are two reasons for thinking (12) does not imply (14*).

First, recall that (14*) is equivalent to

(14**) $P \mathbin{\&} \forall s \forall x \{Fsx \supset \Box((x \mathbin{\&} H) \mathbin{\&} P)\}$.

Now, since (12) is supposed to imply (14*) and since (14*) implies (14**), then (12) should imply (14**). So, we should be able to add any proposition consistently to (12) and (14**) should still be true. But this is not the case.

Argument: suppose, for *conditional proof*, it is true that

(12) P is true and there's nothing anyone is free to do in the circumstances that even might result in ~P.

To simplify things, suppose that P is necessarily true. This is obviously consistent with (12). Now suppose that there is some agent/proposition pair such that the agent is free to make it the case that some proposition is true. That is, suppose that

(16) $\exists s \exists x (Fsx)$.

(16) is obviously consistent with (12). (12) does not preclude someone's being free to make something the case. Moreover, if we know anything about free will, we know that, possibly, someone is free to make a non-necessary (i.e., contingent) truth the case—indeed, it is tempting to think that no one is free to make a necessary truth the case. So, we may consistently suppose with (12) that

---

[44] Warfield (2000: 179).

(17) $\exists x \exists s \{Fsx \ \& \sim\Box x\}$.

However, (12) and (17) are inconsistent with (14**). For *reductio*, suppose that (14**) is true. Suppose that (12) and (17) hold. Then, derive (*Fab* & ~$\Box a$) by instantiating (17), from which it follows that ~$\Box a$. From *Fab* and (14**), derive $\Box((a \ \& \ H) \ \& \ P)$. From $\Box((a \ \& \ H) \ \& \ P)$, derive $\Box a$. Thus, ($\Box a$ & ~$\Box a$). Thus, (14**) is false, which completes the *reductio*. Thus, if (12) is true, then (14**) is false, which completes the *conditional proof*. Thus, (14**), and so too (14*), incorrectly translates (12).

　　　　The second reason for thinking that (14*) incorrectly translates (12) consists of an argument parallel to the one just given. The idea is that it is possible that (12) is true, an agent is free to make something the case, but the agent's circumstances are not broadly logically necessary. For example, an agent need not be a necessarily existent entity, yet the circumstances in which an agent acts include that the agent exists. Hence, the circumstances are not necessary. In fine, free to make something the case does not require that the past itself be broadly logically necessary. Thus, supplant (17) with

(18) $\exists s \exists x \{Fsx \ \& \sim\Box H\}$

and run through the previous conditional proof accordingly. Consequently, (12) does not entail (14**). Thus, there are two good arguments for thinking that (14**), and so too N&R's (14*), incorrectly translates (12). Hence, N&R's rejection of Warfield's argument for (1) fails.

3.4.2.2   Tom Flint's Objection
　　　　In an unpublished piece, Thomas Flint (1998b) argues that no reasonable compatibilist should find Warfield's argument for (1) convincing. Like N&R, Flint contests Warfield's symbolic translations of the natural language expressions captured in (12) and (13). Flint's criticism is two-tiered. First, unlike N&R, Flint provides an argument for thinking that one of Warfield's translations is mistaken. Second, like N&R, Flint deploys his own translations of (12) and (13), illustrating how the target inference from (12) to (13) goes through without obviously requiring (1). But unlike N&R, Flint's translations do not obviously fall prey to counterexample. I'll briefly rehearse Flint's damaging objections.
　　　　So, what is Flint's argument for thinking that Warfield's translations are incorrect? Flint reasons that if Warfield is right that

(15) $P \ \& \sim\exists s \exists x \{Fsx \ \& \ \Box((x \ \& \ H) \supset \sim P)\}$

correctly translates

(13)   *P* is true and there's nothing anyone is free *to do in the circumstances* that would definitely result in ~*P*,

then

(44) $P \ \& \ \exists s \exists x \{Fsx \ \& \ \Box((x \ \& \ H) \supset \sim P)\}$

correctly translates

(45)   P is true and there is something someone is free to do in the circumstances that would definitely result in ~P.

However, no seasoned compatibilist should accept (44) as the correct translation of (45). Hence, no seasoned compatibilist should accept (15) as the correct translation of (13).[45]

So, why does Flint think that no compatibilist should accept (44) as the translation of (45)? Because a reasonable compatibilist can construct a case where (44) is true while (45) is false. Flint (1998b: 9) provides the following instance of (45), letting 'P' denote a necessary truth:

(45*)   Four is an even number and there is something someone is free to do in the circumstances that would definitely result in four's not being an even number.

Now (45*) is obviously false, by anyone's lights. Indeed, it is necessarily false. The question, then, is whether any reasonable compatibilist can construct a case where the relevant instance of (44) is true. That is, could a compatibilist believe

(44*)   (Four is even) & $\exists s \exists x \{Fsx$ & $\Box((x$ & $H) \supset$ ~(Four is even))\}?

The answer is Yes.[46] A compatibilist may simply imagine a determined agent who is free to do something (call it x) other than what the agent actually does. Of course, from a compatibilist's point of view, were the agent to do x, then the circumstances in which the agent does x would have had to have been different than H. But nothing in (44*) prohibits the compatibilist from this assumption. So, the agent is actually determined not to do x, even though the agent is free to do x. And since x is incompatible with H, the conjunction of x and H entails anything whatsoever, including four's not being an even number. So, we have a case where (44*) is true.

It follows that there is a case where (44*) is true while (45*) is false. Thus, (44*) incorrectly translates (45*). Thus, by parity of reasoning, (15) incorrectly translates (13). So, even if (14) implies (15) only if (1) is true, since (13) fails to translate (15), there's no reason to think that (12) implies (13) only if (1) is true. Therefore, Warfield's argument for (1) fails.

As for Flint's second objection, he offers plausible symbolic translations of (12) and (13), where the target inference goes through without any obvious requirement of (1). Flint (1998b: 11) translates (12) and (13) as

---

[45] The structure of Flint's argument runs as follows. Suppose, for reductio, that Warfield's (15) correctly translates (13). It follows that (13) strictly implies (15). Hence, ~(15) strictly implies ~(13). Nevertheless, a reasonable compatibilist has a counterexample to this last claim. That is, she may entertain a case where (13) is true while (15) is false. (45*) is an instance where (15) is false, and to make (13) true just imagine a case where a determined agent could have freely done something other than what she did do. Interestingly, a similar argument *cannot* be applied against Warfield's translating (12) as (14).

[46] See Flint (1998b: 10).

66

(12*) $P \, \& \, \forall s \forall x \{Fsx \supset (x \,\square\!\!\rightarrow P)\}$

and

(13*) $P \, \& \, \forall s \forall x \{Fsx \supset (x \,\lozenge\!\!\rightarrow P)\}.$[47]

The move from (12*) to (13*) does not obviously employ (1). The inference is easily justified by the standard inference from $(p \,\square\!\!\rightarrow q)$ to $(p \,\lozenge\!\!\rightarrow q)$.

It is worth noting that Warfield anticipates Flint's proposal to translate (12) and (13) with counterfactuals. He (2000: 179 note 17) claims:

> One might want to understand [12] and [13] using subjunctives rather than strict conditionals. I think that this would be a mistake. One preferring such a reading however could construct an argument parallel to the one in the text reaching the same conclusion. I leave the task of constructing the argument to those attracted to the subjunctive interpretation of [12] and [13].

So, why would utilizing subjunctives be a mistake, if the same conclusion follows regardless? Warfield does not say. Moreover, the compatibilist has reason for thinking that the same conclusion does not obviously follow.

In fine, Nelkin & Rickless object to Warfield's two-step argument for incompatibilism at each step. I have shown that neither of their objections is damaging. Nevertheless, I find the power of both of Flint's objections so severe that I see no way to mend Warfield's argument.

3.5   Conclusion

After getting clearer about which sort of determinism philosophers argue is or is not compatible with free will, I isolated the point where incompatibilists believe that determinism needs to break down in order to leave room for free will (provided that free will is possible). I noted that incompatibilism is the lay person's pre-theoretic answer to the Compatibility Question. Though there are plenty of arguments against compatibilism, it's not at all clear (to me, at least) that any of them should convince the reasonable compatibilist.

So, where does one go from here? I find insightful Flint's (1987: 440) reflections:

> Commendable as the search for such an argument [refuting compatibilism] surely is, it may well be that the incompatibilist's best means of fostering philosophical progress lies in a different direction—that of developing a more finely articulated

---

[47] More intuitively, consider the following pair of candidate translations for (12) and (13):

(46) $P \, \& \, \sim\!\exists s \exists x \{Fsx \, \& \, H \, \& \, (x \,\lozenge\!\!\rightarrow \sim\!P)\}$, and

(47) $P \, \& \, \sim\!\exists s \exists x \{Fsx \, \& \, H \, \& \, (x \,\square\!\!\rightarrow \sim\!P)\}$, which are, respectively, equivalent to

(46*) $P \, \& \, \forall s \forall x \{(Fsx \, \& \, H) \supset (x \,\square\!\!\rightarrow P)\}$, and

(47*) $P \, \& \, \forall s \forall x \{(Fsx \, \& \, H) \supset (x \,\lozenge\!\!\rightarrow P)\}$.

Again, the inference from (46*) to (47*) is justified by inferring a might-counterfactual from its corresponding would-counterfactual. The deduction need not employ (1).

version of incompatibilism and responding to objections that compatibilist's or others raise against it. This more positive approach may lack much of the glamour and excitement that attend the search for a refutation of compatibilism, but I suspect that it will in the end prove more fruitful.

The rest of my book aims to defend such a finely articulated version of undetermined free action, *viz*., the theory of agent-causation. We'll see in later chapters that the best theory of an undetermined free action underwrites in a principled way the folk's belief in incompatibilism. We thus acquire a sort of inference to the best explanation, where agent-causation best explains our pre-philosophical data.

Of course, agent-causation is viable view of an undetermined free action only if an undetermined free action is possible. The next chapter aims to discredit an argument for the view that an undetermined free action is impossible, thereby removing a substantive objection to my project.

CHAPTER 4

ON THE POSSIBILITY OF UNDETERMINED FREE ACTION

We imagine we are free in our actions, just as in dreaming we deem a place
perfectly familiar which we then see doubtless for the first time.[1]
— Georg Christoph Lichtenberg

4.1   Peter van Inwagen's Pickle

Peter van Inwagen advances the literature on free will as much as any other leading metaphysician. As we saw in Chapter 3, he (1983, 1975, 1974) crisply defines the thesis of determinism. He (1983, 1975) also provides one of the most rigorous arguments for incompatibilism, the so-called Consequence Argument.

Van Inwagen (2001, 2000) has become increasingly more insistent, however, that there are good arguments for thinking that free will is also incompatible with *in*determinism. So, if all of his arguments succeed, free will would be self-contradictory. Regarding certain arguments concluding that an undetermined free action is impossible, van Inwagen (2001: 19) reports autobiographically, "But as to the latter class of arguments, he's damned if he knows what the flaws in them might be. He simply hasn't a clue."

But surely a free action is possible. Indeed, van Inwagen holds that a free action undeniably exists. So, caught in a philosophical pickle, he concludes that a free action is a mystery.

Van Inwagen (2001: 22) laments that the majority of his fellow incompatibilists do not really appreciate the depth and difficulty of this mystery. Fortunately, van Inwagen does not simply stipulate that there are careful arguments for the incompatibility of free will and *in*determinism. He (2001: 18-25, 2000: 10-19) also advances one of his own, dubbing it the *Mere Matter of Chance Argument* (hereafter, *MMCA*).

This chapter critically assesses van Inwagen's *Mere Matter of Chance Argument* (MMCA), finding it woefully inadequate. I divide the chapter into six sections. §4.1 reviews a few crucial concepts and outlines why many compatibilists as well as most incompatibilists should from the outset find Inwagen's MMCA troubling. I do a bit of meta-philosophy, by circumscribing my intended audience and proposing a standard of measure for assessing which side of the dispute is more credible. §4.2 carefully articulates van Inwagen's MMCA. The remaining four sections, §4.3 - §4.6, undermine the MMCA. §4.3 advances plausible reasons for thinking that van Inwagen's foundational thought-experiment is implicitly incoherent. §4.4 isolates a pivotal premise of the MMCA, arguing both that van Inwagen provides insufficient grounds for this premise and that there are good reasons for rejecting it. §4.5 contends that the MMCA is invalid, as there is a parallel argument that vindicates the possibility of an undetermined free action. Finally, §4.6 proposes a direct argument against van Inwagen's

---

[1] Quoted in A.J. Ayer and Jane O'Grady (1992: 62).

MMCA. If I'm right, there is one less barrier to rendering intelligible the notion of an undetermined free action.

It is worth noting that van Inwagen's MMCA is *not* another instance of the so-called *Mind* argument(s). Rather, it is a refreshingly new argument for the impossibility of an undetermined free action. Moreover, while only two philosophers have contested van Inwagen's MMCA in print, my strategy differs radically from theirs.[2] Finally, studying van Inwagen's argument is fun. He begins with an accessible thought-experiment, and the rest of his argument does not require a familiarity with current discussions about free will.

4.1.1 Defining *Free Will* and *Determinism*: a Review

Let *free will* (FW) designate the thesis that someone freely performs some action. I circumscribed the relevant notion of freedom in Chapter 2. Nonetheless, if my treatment there was somehow inadequate, we may stipulatively understand the concept of free action in the same way that nearly every contemporary expert writing on free will in the analytic tradition does.[3] Intuitively, an agent *freely* performs action A just when she performs A and A is genuinely up to her.

Intuitively, determinism (D) is "the thesis that there is at any instant exactly one physically possible future."[4] More precisely, let *determinism* be the thesis that there is a sufficient causal condition for every event. Event C is a *sufficient causal condition* for event E *iff* C occurs, C does not imply E, and it is a law of nature that if C occurs at a time T, then E occurs at or later than C. Say that an event E is *determined iff* there is a sufficient causal condition for E.[5] Say that an event E is *undetermined iff* E is not determined. It is very important to note that an undetermined event could still be caused. For example, it could be that some event C indeterministically or probabilistically produces event E.[6]

4.1.2 Possible Positions

Why do I contend that many incompatibilists and compatibilists alike should find van Inwagen's MMCA troubling? Consider the following three propositions:

(1) $\Diamond$FW

(2) $\Diamond$(FW & D)

(3) $\Diamond$(FW & ~D)

---

[2] See Randolph Clarke (2003: 164-8) and Laura Ekstrom (2003).

[3] Experts such as Randy Clarke (2003), John Fischer (1986b), Carl Ginet (1990), Robert Kane (1996a), Timothy O'Connor (2000a, 1996, 1995), van Inwagen (1983), Gary Watson (1987, 1982), et. al.

[4] Van Inwagen (1983: 3), who in turn cites Sidney Hook, ed. (1958).

[5] For a more sophisticated definition of a determined event, see Chapter 3's §3.2.

[6] For more in support of this critical distinction, see Anscombe (1998), van Inwagen (1983), Kane (1996a), Clarke (2003), O'Connor (2000a), and Alfred Mele (1995).

Incompatibilists deny (2), claiming that every free action is undetermined. Let's divide the incompatibilists into two groups, the Conventionals and the Radicals. *Radicals* deny (1), holding that a free action is simply impossible. This is a radical position indeed. For, intuitively, someone's being morally blameworthy for something guarantees that someone might have acted freely.[7] And it is possible that someone is morally blameworthy for something. Hence, (1) is true. More precisely, Radicals confront the following argument head-on:

(4)   □{(Someone is morally blameworthy for something) ⊃ ◊FW}[8]

(5)   ◊(Someone is morally blameworthy for something)

_____

(1)   Thus, ◊FW

Defending this argument is not my aim. Instead, I deploy it. I register my conviction that (4) and (5) are self-evident. Radicals, I take it, have their work cut out for them.

In contrast to Radicals, *Conventionals* are incompatibilists who accept (1), which commits them to

(6)   ◊(There is an undetermined free action).

(6) clearly implies (3). The vast majority of incompatibilists are Conventionals. *Libertarians* are Conventionals who also believe that someone actually acts freely.

(2), of course, designates the thesis of compatibilism. Let's divide the compatibilists into three types, depending on their posture toward (3): the Super Inflexibles, the Inflexibles, and the Flexibles.

Super Inflexibles are compatibilists who hold that, necessarily, a free act occurs only if every event is determined. So, they reject (3), which commits them to rejecting (6). However, suppose that a free determined act occurs. If this is possible, then surely we may consistently suppose that an undetermined event occurs somewhere else, perhaps in the corner of a distant galaxy and perhaps even later than the determined free action. One determined event does not imply the thesis of determinism, which is a thesis concerning all events. Therefore, Super Inflexibles are mistaken.

Inflexibles are compatibilists who maintain that, necessarily, every free action is determined. So they hold that only determined actions could count as free.[9] Inflexibles may accept (3). For example, Inflexibles may suppose that there is a determined free action while

_____

[7] See Chapter 2, §2.3.1.

[8] Notice that (4) is implied by the common claim that, necessarily, an agent is morally blameworthy only if this agent acts freely. But, despite its popularity, this common claim is not obviously true. Suppose that an agent never acts. Still, she might be morally to blame for not acting as she could have acted. However, had she acted in the way she ought to have acted, it seems clear that she would have acted freely. Thus, she might have acted freely. Hence, ◊FW. Therefore, we have a plausible counterexample to the common claim that nonetheless confirms (4).

[9] Arguably, David Hume (1955) and Thomas Hobbes (1969) are Inflexibles.

somewhere else there occurs an undetermined event, in which case they may reasonably believe (3). So, while Inflexibles may or may not accept (3), they clearly deny (6). Nevertheless, even if compatibilism is right, Inflexibilism arguably confronts problems.

To see why, consider a possible world W in which event C determines that Sam Slug performs free action A. That is, C is a sufficient causal condition for Sam Slug's freely performing A. Now consider a possible world W* that is exactly like W except that the laws of nature are just a touch different. In W*, there is no law of nature that if C occurs at time T, then Sam Slug freely performs A at or later than C. Instead, there is a very similar law of nature. That is, in W* there is a law of nature that if C occurs at time T, then it is 99.999% likely that Sam Slug freely performs A at or later than C. However, we may consistently suppose that C still causally produces Sam Slug's action. Moreover, it seems clear that if Sam Slug acts freely in W, then Sam Slug acts freely in W*. For, *the very same events* cause his action in both worlds. His action's being just-a-tiny-touch-less-than-determined in W* should not suffice for his action's being unfree. So, it seems clear that Sam Slug's free action A is undetermined in W*.

The crucial claim is that if there is a determined free action in one world, then were the same action to occur in virtue of being very nearly causally determined but still caused, then this undetermined action would be free as well. More precisely, the following proposition seems true:

(7)   □{(Event C determines someone's freely performing act A) ⊃ ◊(C very nearly determines yet still fully causes someone's freely performing act A)}.[10]

Since (7) is true, Inflexibilism is wrong. The moral is that Inflexibles ought to be more flexible, as there might be an undetermined free action. They should give up Inflexibilism, embracing (6) instead.

Flexibles, then, are compatibilists who believe (6), i.e., there might be an undetermined free action. Flexibles, then, believe that there might be determined free actions *and* that there might be undetermined free actions. That is, Flexibles hold to both (2) and (6), which commits them to (3). I think that the vast majority of compatibilists should be Flexibles.[11] Flexibilism is *prima facie* the most reasonable compatibilist position, and it is even consistent with the claim that every free action is actually determined.

4.1.3   My Intended Audience and a Standard of Measure
So, both Conventionals and Flexibles accept (6). But (6) is just the denial of the MMCA's conclusion. Thus, the MMCA succeeds only if both Flexibles and Conventionals are mistaken. Therefore, van Inwagen's MMCA succeeds only if many compatibilists and the vast majority of incompatibilists are mistaken. So, who is right? Conventional incompatibilists and Flexible compatibilists, on the one hand? Or Peter van Inwagen, on the other hand?

---

[10] Notice that incompatibilists are committed to (7), if only for the reason that the antecedent is necessarily false.

[11] See Mele (1995) and William Lycan (2003).

My intended audience does not include van Inwagen.[12] *He has already made up his mind. Nor does my intended audience include Conventionals and Flexibles. They too have already made up their minds. Rather, I write so as to convince what I'll call the Jury.* The Jury *consists of those intelligent, reflective, and reasonable folk who do not have the issue settled in their minds.*

There is a standard of measure for assessing which side of the dispute is more credible. Consider (1) and (7) once again. They are both very intuitive. Notice that they together imply (6). Proof: suppose that (1) and (7) are true. So, there might be a free action A. Consider the world including free action A. Now, either A is determined or not. Take the first horn, supposing that A is determined. It follows from (7) that there might be an undetermined free action. Hence, (6) is true. Take the second horn, supposing that A is not determined. Since A is free, it follows that A is free and undetermined. Hence, (6) is true. Therefore, regardless of whether A is determined, (6) is true. Q.E.D.

Recall the two purportedly self-evident premises that strictly imply (1), *viz.*, (4) and (5). Thus, the following proposition has an overwhelmingly great deal of intuitive appeal:

(8)   Both (1) and (7) are true.

Moreover, many compatibilists (i.e., the Flexibles) and most incompatibilists (i.e., the Conventionals) accept (8). And since (8) implies the falsity of (6), (8) implies that van Inwagen's MMCA is unsound. Our standard for the Jury, then, is this:

(Standard)   One's degree of confidence that van Inwagen's MMCA succeeds (or fails) should vary proportionally with one's degree of confidence that its validity and basic premises are more (or less) intuitive than (8).

Consequently, if one finds (8) more intuitive than at least one of the MMCA's premises, then one should deem the MMCA a failure. Moreover, if one finds (8) more intuitive than accepting the validity of van Inwagen's MMCA, then one should reject the MMCA.

In what follows, I argue that both Flexibles and Conventionals may reasonably reject *both* the MMCA's validity as well as *several* of its basic premises. For, they may stand unashamed in their commitment to (8), which upon reflection is far more tenable than the soundness of van Inwagen's MMCA. I therefore sympathize with and attempt to bear out Thomas Reid's (1895: 337) conviction:

> Every argument in a dispute, which is not grounded on principles granted by both parties, is that kind of sophism which logicians call *petitio principii*; and such, in my apprehension, are all the arguments offered to prove that liberty [i.e., freedom] of action is impossible.

If I am right, van Inwagen's argument need not oblige every compatibilist and incompatibilist to give up his view. Let the Jury decide.

---

[12] Though I thank him deeply for agreeing to be one of my book's readers, I can only pretend to proffer an argument he would find convincing.

4.2   Van Inwagen's *Mere Matter of Chance Argument*[13]

Van Inwagen's MMCA is designed to convince any reasonable person that an undetermined free action is impossible. Van Inwagen contends that while his argument is "very informal" it is nonetheless "intuitive" and "has a great deal of plausibility." He (2000: 11) aims to show that any occurrence of an undetermined act is a mere matter of chance and therefore the act cannot be *up to* the agent who performs it.

It may be worth noting that van Inwagen (1983: 128-9) once argued that no good argument should rely on 'chance' or 'random' as a load-bearing term. Recently, he (2000: 13) recalls, "I [once] argued there that the words 'random' and 'chance' most naturally applied to *patterns* or *sequences* of events, and that it was therefore not clear what these words could mean if they were applied to single events [such as the occurrence of a free action]." However, van Inwagen has since changed his mind. He (2000: 13) now prefaces his MMCA by stating, "It will be evident from what follows that I no longer regard this [no-one-knows-what-'chance'-means-] argument as having any merit."

For ease of explication, I divide his argument into the following four phases: setup-phase, the single-replay-phase, the multiple-replays-phase, and the impressions-phase.

4.2.1   The Setup-Phase

The argument starts with a story designed to describe the clearest case of an agent's acting freely. For a working example, imagine that Alice was faced with a difficult choice between lying and telling the truth. Suppose that it was determined that *either* she lied or she told the truth, i.e., that she makes a choice is forced. However, we assume *for reductio* the following proposition:

(Reductio)   Alice's original free action was undetermined.

So, it was not determined that she lied, nor was it determined that she told the truth. Suppose, though, that as things turned out, she freely told the truth. That is, suppose that Alice seriously considered telling the truth, she seriously considered telling the lie, she was able to tell the truth, and she was able to tell the lie she had been contemplating. Let "α" designate the world where Alice actually (freely) tells the truth. Again, it follows from Reductio that Alice's telling the truth, though purportedly a free act, was undetermined.

4.2.2   The Single-Replay-Phase

Now suppose that immediately after Alice told the truth, God caused the universe to revert to precisely its state one minute before Alice told the truth and then let things "go forward again," where T1 is the moment one minute before she (freely) told the truth, and T2 is the second time the universe is in this state. Van Inwagen (2000: 14) asks, "What would have happened the second time? What would have happened after T2? Would she have lied or would she have told the truth?"

---

[13] Van Inwagen (2001: 18-25, 2000: 10-19).

Van Inwagen answers that there simply is no fact as to what Alice *would* do. For, he contends that, necessarily, if her "original" decision to tell the truth one minute after T1 was undetermined by the state S1 of the world at T1, then her decision about a minute after T2 was undetermined by the state S2 of the world at T2. And since S2 does not determine which action Alice performs about a minute after T2, van Inwagen (2000: 14) asserts that "one can say only that she *might* have lied and she *might* have told the truth."

It follows, according to van Inwagen, that were S2 the state of the world at T2, then it might be the case that Alice tells the truth about a minute later—stipulatively read as "(S2 ◊→ Truth)." Likewise, were S2 the state of the world at T2, then it might be the case that Alice lies about a minute later—stipulatively read as "(S2 ◊→ Lie)." As Lewis (1973) notes, it is a theorem of standard counterfactual semantics that, necessarily, for any two contingent propositions P and Q, (P ◊→ Q) *iff* ~(P □→ ~Q). Thus, from (S2 ◊→ Truth), infer ~(S2 □→ ~Truth). And from (S2 ◊→ Lie), infer ~(S2 □→ ~Lie).

Recall, though, that her having to make exactly one of the choices was forced. She had to either lie or tell the truth. So, from ~(S2 □→ ~Truth) and ~(S2 □→ ~Lie), it follows that it was neither the case that Alice would have told the truth given the circumstances S2 nor the case that Alice would have lied given the circumstances S2. Therefore, since her act was undetermined by the circumstances, there was no fact of the matter as to what she *would* have done in the circumstances.

4.2.3 The Multiple-Replays-Phase

We then imagine that God puts Alice through this ordeal again and again and again, *ad nauseum*. We (2000: 14) suppose that "God a thousand times caused the universe to revert to exactly the state it was in at T1 (and let us suppose that we are somehow suitably placed, metaphysically speaking, to observe the whole sequence of 'replays')." For the same line of thought given in the Single-Replays-Phase, van Inwagen argues that for *each* replay there is no fact of the matter as to which act (of the two) Alice would have performed. It follows that there is no fact of the matter as to how the series of a large number of replays would unfold prior to its unfolding. Rather, since it is most certainly the case that she would probably not decide the same way each time, we should only expect to have observed that sometimes Alice lied and sometimes she told the truth. So, were God to replay the incident a ludicrously large number of times, the ratio of the outcome "truth" to the number of replays would converge on, would settle down to, some number less than one. Perhaps in thirty percent of the replays Alice told the truth and in the remaining 70% of the replays she lied. However, van Inwagen requests that we imagine the simplest case, where the ratio of the outcome "truth" to the number of replays was ½. Thus, the ratio of the outcome "lie" to the number of replays was also ½. Call such a world a *flip-flop-world*.

4.2.4 The Impressions-Phase

Van Inwagen (2000: 15) asserts,

> If we have watched seven hundred and twenty-six replays, we shall be faced with the inescapable impression that what happens in the seven-hundred-and-twenty-seventh replay will be due simply to chance.

He claims that there is no reason for resisting this impression as there is nothing else we could learn about the situation that could undermine the impression that what happens in the *next* replay will be a mere matter of chance. For, van Inwagen reasons, we already know everything that is relevant to evaluating whether or not the next replay will be a mere matter of chance.

Most important, the relevant knowledge van Inwagen has in mind consists of this. The laws of nature and the state S728 of the world will not together guarantee the outcome of the 727th replay at time T728.[14] Van Inwagen (2000: 15) explains, "Each time God places the universe in this state, both 'truth' and 'lie' are consistent with the universe's being in this state and the laws of nature." Given what we observed in the multiple replays, van Inwagen (2000: 15) reasons thus:

> A sheaf of possible futures (possible in the sense of being consistent with the laws) leads "away" from this state, and, if the sheaf is assigned a measure of 1, surely, we must assign a measure of 0.5 to the largest sub-sheaf in all of whose members Alice tells the truth and the same measure to the largest sub-sheaf in all of whose members she lies. We must make this assignment because it is the only reasonable explanation of the observed approximate equality of the "truth" and "lie" outcomes in the series of replays. And if we accept this general conclusion, what other conclusion can we accept about the seven-hundred-and-twenty-seventh replay (which is about to commence) than this: each of the two possible outcomes of this replay has an objective, "ground-floor" probability of 0.5—and there's nothing more to be said? And this, surely, means that, in the strictest sense imaginable, the outcome of the replay will be a matter of chance.

So van Inwagen thinks that reasonable people would be compelled by their properly functioning epistemic faculties to think that the outcome of the 727th replay will be due simply to chance. Moreover, since we are reasonable people, once we have the inescapable impression that what happens in the seven-hundred-and-twenty-seventh replay will be due simply to chance, we are likewise compelled by our rationality to have the conviction that what happened on any replay was due simply to chance. Indeed, the unfolding of the initial segment of each replay, by design, was exactly like the unfolding of the initial segment of the original course of events before God miraculously intervened. Thus, since there is no relevant difference between the replays and the original situation, we again are bound by our rationality to think that Alice's original act of telling the truth was due simply to chance. Van Inwagen (2000: 15) stresses:

> Now, obviously, what holds for the seven-hundred-and-twenty-seventh replay holds for all of them, including the one that wasn't strictly a *re*play, the initial sequence of events. But this result concerning the "initial replay," the "play," so to speak, should hold whether or not God bothers to produce any replays. And if He does not—well, that's just the actual situation. Therefore, an undetermined

---

[14] Where "S728" designates the state the world is in at the beginning of the 727th replay, just like "S2" designates the state of the world at the beginning of the first replay. Presumably, S728 is the 728th time the state of world has been in precisely the state S1 it was in one minute before Alice first told the truth.

action is simply a matter of chance: if it was undetermined in the one, actual case whether Alice lied or told the truth, it was a mere matter of chance whether she lied or told the truth.

Finally, van Inwagen takes it as a conceptual truth that no one is able to determine the outcome of a process that is due simply to chance. That is, necessarily, if the outcome of a process is due simply to chance, then the outcome is not up to anyone. It follows that if it is a mere matter of chance whether Alice will lie or tell the truth, then Alice has no free will with respect to lying or telling the truth. If which course of action Alice takes is a mere matter of chance, then Alice is not *able* to opt for exactly one course of action. Thus, since we are only reasonable in thinking that Alice's action was simply due to chance, we are only reasonable in thinking that Alice did not freely tell the truth, contrary to our original assumption. Presumably, this conclusion can be generalized to cover any agent's undetermined action in any world. Thus, a free undetermined action is impossible.

Now we can see why van Inwagen faults his earlier argument for the conclusion that no argument relying on the terms 'chance' or 'random' should be taken seriously. Van Inwagen now denies the *inference* he once accepted. That is, he denies the move from (i) 'chance' most naturally applies to patterns or sequences of events to (ii) it is unclear what the term could mean when applied to a single event. It is unclear whether or not van Inwagen still believes that 'chance' most naturally applies to patterns or sequences of events. However, his MMCA suggests that one can use the conjecture that 'chance' applies to patterns or sequences of events to illuminate what the term 'chance' means when applied to a single event. For, the philosopher purportedly can take any single process that issues in a single event and focus on the pattern or sequence of events that would result from repeating the initial conditions of *that* process a sufficiently large number of times. Then, seeing that the overall pattern or sequence of outcomes of those processes is exactly what one should expect were each outcome in the sequence randomly generated, one should, in the absence of defeating reasons, believe that each outcome is due simply to chance.

4.3   First Objection: On the Possibility of Replays
        I dispute nothing in the setup phase. However, replaying the situation according to van Inwagen's stipulations may be impossible. Perhaps it is possible. However, it's simply too controversial a premise on which to rest the substantive conclusion that both Flexibles and Conventionals are mistaken in thinking that there might be an undetermined free action. Here are three reasons why God's replaying the situation—again, according to van Inwagen's stipulations—may be impossible.
        First, it may be that Alice could not survive God's replaying the incident. Philosophical problems of personal identity are complicated and thorny. Whatever the received view of personal identity, reversing natural causal processes or abrupt shifts in causal processes or leaps in causal processes can only diminish the intuition that the agent persists through the reversal, shift, or leap. It seems reasonable to believe that persisting objects, including agents like Alice, require that there is a causal connection of some sort between the agent at one time and the agent at a later time. It is not unreasonable to suspend judgment on whether van Inwagen's argument works until one sees a plausible account that characterizes how the causal transaction from one replay to the next replay works. I register hesitation regarding the claim that a world where God

replays the incident according to van Inwagen's specifications presents no special problems for plausible theories of personal identity. I conclude that at the very least there are two claims the Jury should compare with respect to their levels of intuitive strength. First, Alice could survive such a replay. Second, (8) is true. The first claim, if true, is not as obviously true as the second claim. However, a good argument against the possibility of a free undetermined action should contain premises all of which are more intuitively true than the claim that (8) is true. So, by Standard, the Jury should deem van Inwagen's argument unconvincing.

Second, van Inwagen's stipulations implicitly assume that the laws of nature in the flip-flop-world are the same as the laws in α, where α is the actual world where, by Reductio, Alice freely told the truth. But the two worlds' sharing precisely the same laws may be impossible. The laws in the flip-flop-world are consistent with God's replaying a certain course of events, yet the laws in α are not obviously consistent with God's replaying these events. However, at least Conventionals (e.g., libertarians) may reasonably hold that a difference in the laws of nature undermines any relevant comparison between Alice's act in α with Alice's act in the flip-flop-world. After all, Conventionals take seriously the claim that an agent's freely acting requires that the agent be able to refrain from so acting given the exact same laws of nature. I conclude that at the very least the Jury has two claims to compare with respect to their levels of intuitive strength. First, the laws of nature in the nearby flip-flop-world are the same as the laws in α. Second, (8) is true. The first claim, if true, is not as obviously true as the second claim. So, by Standard, the Jury should find van Inwagen's argument unconvincing.

Third, contrary to van Inwagen's assumption, it is impossible that Alice acts in the same circumstances in each replay. More precisely, Alice cannot act in precisely the same complete nondetermining circumstances in each replay. Say that C is a set of *complete circumstances* of an agent's act just if C includes the entire causal history of the world prior to the time of the agent's act, together with whatever outside influences are affecting the agent at the time of her action.[15]

To illustrate, let "C1" designate the set of complete circumstances in which Alice tells the truth the very first time. Now suppose that God causes the universe to revert to precisely the state it was in one minute before Alice told the truth in C1 and then lets things "go forward again." Alice either lies or tells the truth within the next few minutes. Suppose that she then lies. Given Reductio, Alice lies in some set C2 of complete nondetermining circumstances. But notice that C2 includes C1 and more besides. For example, C2 includes *God's causing the universe to revert to precisely the state it was in one minute before Alice told the truth in C1*. Thus, C2 includes *Alice's telling the truth in C1*. Of course C1 does not include the state of affairs *Alice's telling the truth in C1*, since C1 is according to Reductio nondetermining with respect to which act Alice performs. Thus, C2 ≠ C1. Moreover, these additional circumstances (besides C1) that C2 includes are part of the etiology of Alice's telling the lie. For example, the event *God's causing the universe to revert to precisely the state it was in one minute before Alice told the truth in C1* contributes causally to precisely which state the universe is in during the next few moments, which accordingly contributes causally to Alice's telling the lie roughly one minute later. Thus, since the relation of causal contribution relation is obviously transitive, the causal contributors of Alice's telling the lie at the end of the first replay are different than the causal contributors of Alice's telling the truth the first go around. In other words, C2 is distinct

---

[15] For more on cashing out an agent's complete circumstances, see Flint (1998a: 47, 200).

from C1 at least partly in virtue of the difference in the total causal influences on which act Alice performs. Generalizing, we may infer that it is impossible that Alice is in the same complete circumstances for each replay.

Since Alice is in different complete circumstances each time she is in a position either to lie or to tell the truth, the Jury should wonder whether this difference in etiology explains why her not acting the same way every time is due simply to chance.[16] At the very least, (8) is more intuitive than thinking that the result of each replay is a mere matter of chance even though no replay shares its etiology with any other replay. Given Standard, the Jury should have little confidence in van Inwagen's MMCA.

Let's suppose, for sake of argument, that Alice could survive God's replaying the scenario, that the laws of nature in the flip-flop-world either are the same as the laws in α or else are sufficiently relevantly similar, and that *for all practical purposes* the same set of causal factors influencing Alice's act hold for each replay. In other words, let's just assume that the above objections *per se* are insufficient to undermine van Inwagen's argument. However, even granting that God's replaying the scenario is possible, problems remain.

4.4   Second Objection: On Counterfactuals of Undetermined Events
Van Inwagen stresses that for every time Alice is in the position either to lie or to tell the truth, there can be no fact of the matter as to which act Alice *would* perform. That is, van Inwagen believes:

(9)   Necessarily, for any agent Δ who performs any action A in any set C of complete nondetermining circumstances, it is false that *were Δ in C, then Δ would perform A*. To abbreviate, "$(C \& A) \Rightarrow \sim(C \,\square\!\!\rightarrow A)$."[17]

This section addresses two questions. First, what role does (9) play in van Inwagen's MMCA? Second, should every reasonable thinker accept the reasons van Inwagen offers for (9)? I discuss how van Inwagen's thought-experiment would be undermined were (9) false. I then argue that reasonable Flexibles and Conventionals need not accept van Inwagen's case for (9). The Jury, then, should think that (9) is more controversial than (8). Hence, by Standard, the Jury should reject the MMCA.

4.4.1   On the Role of Proposition (9)
So, what role does (9) play in van Inwagen's MMCA? There are at least two candidates. First, (9) carries a sort of rhetorical momentum in favor of freedom-squashing chance. Some

---

[16] Many philosophers believe that an event's causal history is essential to the event. They believe that, necessarily, if event E occurs and has a causal history H, then it is necessarily the case that E occurs only if H is E's causal history. If they are right, then Alice could not on two separate occasions perform the same action. I am not one of these philosophers.

[17] Elsewhere, van Inwagen (1997) argues that $\sim\lozenge(C \,\square\!\!\rightarrow A)$. That is, van Inwagen (1997) holds that it is broadly logically impossible that some agent *would* perform some action in complete circumstances nondetermining with respect to the action performed. If this is right, then (9) is true in virtue of a necessary consequent.

propositions have that familiar "ring of truth." They are intuitive. Analogously, (9) has the "ring of chance." For, (9) implies that even though Alice tells the truth in the circumstances, since her telling the truth is undetermined by

the circumstances, there *just is no fact of the matter* as to what she *would* do in the circumstances. If there is no act Alice *would* do in the circumstances, her telling the truth rather than lying seems chancy. Not even an omniscient being could know in advance what she would do, for there is nothing there to be known.

Alternatively, suppose that after 726 replays God tells us, "For each time Alice was in the position either to tell the truth or to tell the lie, it was the case that, for exactly one of those alternatives *x, were she in that situation at that time, she would opt for alternative x*. Moreover, if I, God, were to replay the situation for the 727th time, she would lie." Perhaps God's assertions could reasonably undermine the impression we might have about whether the next outcome's occurrence will be due simply to chance.

So the falsity of (9), intuitively speaking, silences the "ring of chance" that the description of the scenario may have. Alternatively, the truth of (9) seems to confirm that "ring of chance." Defeating belief in these counterfactuals of undetermined action seems to neutralize any little but nagging hunch one may have in favor of the following claim. Since Alice freely acts, there must be a fact of the matter as to what she would do, as there should be something— say, Alice herself—that makes one course of action materialize rather than the alternative course of action.

The second and perhaps more important role of (9) is that it apparently paves the way, dialectically speaking, for taking the flip-flop-world seriously. For, suppose that the flip-flop world is not a nearby world. The libertarian may contend that since the flip-flop-world is not sufficiently similar to α, then what goes on in the flip-flop-world—*viz*., Alice's flipping and flopping between lying and telling the truth—is irrelevant to whether Alice acts freely in α. Indeed, if (9) is false, the libertarian may hold that in normal circumstances, such as Alice's telling the truth in α,[18] it may be that were God to replay the incident a ludicrously large number of times, Alice would do the same thing every single time (in which case we would <u>not</u> acquire the inescapable impression that the outcomes are due simply to chance). Call a world where Alice performs the same undetermined act in every one of a large number of replays a *steady-world*. If steady-worlds exhaust the sufficiently similar worlds that are multiply replayed, then the flip-flop-world is not sufficiently similar and (9) is false.

In fine, if one already believes that (C □→ A), then if God were to replay the situation, Alice would perform the same act again and again. That is, if one already believes that (C □→ A), then steady-worlds exhaust the sufficiently similar worlds to a world including (C & A). Since van Inwagen denies this consequent, he attempts to undermine belief in such counterfactuals of undetermined action with (9).

4.4.2   On van Inwagen's Case for (9): the Might-Argument

As the reader may recall from the single replay phase, §2.2, van Inwagen attempts to provide such a good argument for (9)—call it the *Might-Argument*. We turn now to our second question: Should every reasonable libertarian accept the Might-Argument? I contend that there

---

[18] I assume that God's replaying the situation is not normal.

are a few considerations in support of (9)'s falsity, and there are good grounds to reject the Might-Argument.

Before evaluating the Might-Argument, I shall briefly provide some degree of support for the falsity of (9). First, many reasonable libertarians actually believe—and believe upon careful reflection—the following proposition:

(10)   Necessarily, an agent Δ freely performs action A in a set C of complete nondetermining circumstances only if either (i) if agent Δ were in C, Δ would freely perform A, or (ii) if agent Δ were in C, Δ would refrain from performing A freely.

Now (10) implies the falsity of (9), and yet (10) is theoretically fruitful in ways other than refuting an argument against the possibility of an undetermined free act. For example, (10) lies at the heart of *Molinism*, which is a theory attempting to reconcile God's sovereignty with creaturely freedom. Molinism has tremendous explanatory power.[19] Since (10) both bears magnificent conceptual fruit and implies the falsity of (9), it is reasonable to think that (9) seems far too controversial to be a load-bearing premise in an argument for the impossibility of (undetermined) free action.

If this criticism against (9) is on target, then we should be able to find something wrong with van Inwagen's Might-Argument for (9). We may unpack his argument as follows:

Van Inwagen's Might-Argument

(11)   Alice (freely) tells the truth in set C1 of complete nondetermining circumstances.

(12)   Thus, Alice might tell the truth in C1, and Alice might lie in C1.

(13)   Thus, if it were the case that C1, then it might be the case that Alice tells the truth; and if it were the case that C1, then it might be the case that Alice lies. That is, (C1 ◊→ Truth) & (C1 ◊→ Lie).

(14)   Thus, it is not the case that if it were the case that C1, then Alice would lie; and it is not the case that if it were the case that C1, then Alice would tell the truth. That is, ~(C1 □→ Lie) & ~(C1 □→ Truth).

Van Inwagen justifies the inferences as follows. (12) follows from (11), since the circumstances are *nondetermining*.[20] Since (12) and (13) are equivalent, (12) implies (13). The next inference

_____

    [19] Molinism provides a viable account for explaining the problem of evil, the infallibility of both the Pope speaking ex cathedra and Holy Scriptures with respect to their authorship and canonization, prophecy, unanswered prayers, religious exclusivism, and the incarnation of Jesus Christ. Reasonably minded luminaries holding to Molinism include David Basinger (1987), Rod Bertolet (1993), Eef Dekker (2000), Thomas Flint (2001, 1998a, 1991, 1988) Alfred Freddoso (1988), William Craig (1998, 1994, 1991), J.P. Moreland, Douglas Geivett (2002), Richard Otte (1987), Alvin Plantinga (1985, 1977), Edward Wierenga (2001, 1989), and, aptly enough, Luis de Molina (1988).

    [20] Moreover, since freely performing act A pre-theoretically requires an ability not to perform A freely, there is additional warrant for the move from (11) to (12).

from (13) to (14) relies on a theorem of standard counterfactual semantics, which states that, necessarily, for any two contingent propositions P and Q, (P $\diamond\!\!\rightarrow$ Q) *iff* ~(P $\square\!\!\rightarrow$ ~Q). More precisely, (13) implies that ~(C1 $\square\!\!\rightarrow$ ~Truth) & ~(C1 $\square\!\!\rightarrow$ ~Lie). Since we assume that C1 determines that *either* Alice tells the truth or Alice lies, her not telling the truth implies her lying and her not lying implies her not telling the truth. That is, C1 $\Rightarrow$ (Truth or Lie). A bit of reflection shows that (14) follows directly. That is, given ~(C1 $\square\!\!\rightarrow$ ~Truth) & ~(C1 $\square\!\!\rightarrow$ ~Lie) and C1 $\Rightarrow$ (Truth or Lie), it follows that ~(C1 $\square\!\!\rightarrow$ Lie) & ~(C1 $\square\!\!\rightarrow$ Truth).

   Notice that the Might-Argument need not rely on the claim that the undetermined event is an act. Any undetermined event will do. Thus, (11) may just as well read:

(11*) There is a set C of complete nondetermining circumstances with respect to event E such that E occurs in C.

Where "C" denotes any set of complete nondetermining circumstances and "E" denotes any event, call conditionals of the form "(C $\square\!\!\rightarrow$ E)" *counterfactuals of undetermined events* (*CUE*s).

   So, van Inwagen's Might-Argument is sound only if (11*) entails that there are no CUEs. Suppose, for reductio, that the Might-Argument is sound. Therefore, (11*) entails

(14*) ~(C $\square\!\!\rightarrow$ E) and ~(C $\square\!\!\rightarrow$ ~E).

Generalizing, it follows that

(15) Necessarily, for any set C of complete nondetermining circumstances with respect to event E such that E occurs in C, ~(C $\Rightarrow$ E) only if ~(C $\square\!\!\rightarrow$ E),

which is equivalent to

(16) Necessarily, for any set C of complete nondetermining circumstances with respect to event E such that E occurs in C, (C $\square\!\!\rightarrow$ E) only if (C $\Rightarrow$ E).[21]

Now (16) is a controversial result. It is not obviously true and seems to be false, in which case we should discharge our reductio assumption that van Inwagen's Might-Argument is sound. After all, for propositions x and y, unless y is necessary, one cannot in general infer (x $\Rightarrow$ y) from ((x $\square\!\!\rightarrow$ y) & (x & y)). Indeed, since strict conditionals entail their corresponding counterfactuals, (16) is equivalent to

(17) Necessarily, for any set C of complete nondetermining circumstances with respect to event E such that E occurs in C, (C $\square\!\!\rightarrow$ E) if and only if (C $\Rightarrow$ E).[22]

---

   [21] ~(C $\Rightarrow$ E) is simply stating a consequence of being nondetermining. That is, if C is nondetermining with respect to E, then ~(C $\Rightarrow$ E).

   [22] It might be worth noting what may be an interesting feature of (17). (17) appears to warrant another analysis for the thesis of (causal) determinism. Perhaps we could let 'determinism' denote the conjunction of the following two theses:

But (17) is far from obviously true, yet van Inwagen is committed to thinking that (17) not only follows from (11*) but is logically equivalent to (11*).

Here is a reason for thinking (17) is false. Suppose that you toss a genuinely fair coin twenty times. The odds of the coin's landing heads every time is fantastic (1 in 1,048,576). But it *might* happen. It is very probable, though not inevitable, that the coin does not land heads every time. Surely it is true that were you to toss the coin twenty times, it would *probably* not land heads every time. Suppose that the highly probable happens: it actually does not land heads every time. It *seems* quite intuitive that

(18)   Were you to toss the coin twenty times, the coin would not land heads every time.[23]

It is reasonable to believe (18). Yet (18) strictly implies that there could be true counterfactuals of undetermined events (CUEs). That is, (18) entails the falsity of (17). Since the soundness of van Inwagen's argument implies (17), it is reasonable to think that his argument should fail.

I have built a sort of cumulative critique against van Inwagen's Might-Argument. However, so far I have contested neither the truth of an underived premise nor the argument's validity. Since the only underived premise is (11) and no libertarian should challenge (11), the reasonable libertarian should be able to challenge the argument's validity. I shall argue that the argument is not obviously valid, thus extending my cumulative critique.[24] In particular, I argue that the libertarian should challenge the argument's validity by contesting the move from

(12)   Alice might tell the truth in C1, and Alice might tell the lie in C1

to

- For every instant of time, there is a proposition that expresses the entire state of the world at that instant;
- for laws of nature L and any propositions p and q expressing the entire state of the world at some instants, $(p \ \& \ L) \ \Box \!\!\rightarrow q$.

Needless to say, van Inwagen's (1983) popular *Consequence Argument* would not have near the intuitive appeal under this definition of determinism. Pick your favorite interpretation of the no-choice operator 'N'—e.g., let 'Np' abbreviate 'proposition p is true and no one has, or ever had, any choice about whether p is true'. Van Inwagen (2000: 3-10, esp. 8) now has a new interpretation of 'Np'. Nonetheless, it is far from clear that N$((p \ \& \ L) \ \Box \!\!\rightarrow q)$. And even though it may be intuitive that N$(p \ \& \ L)$, it is far from clear that one may validly infer Nq from both N$((p \ \& \ L) \ \Box \!\!\rightarrow q)$ and N$(p \ \& \ L)$. It seems, then, that the degree to which one thinks this inference is invalid is the degree to which one should think that either van Inwagen should give up (17) and hence his Might-Argument for (9), or else he should give up his Consequence Argument.

[23] Again, the reader may adjust the odds to suit her own intuitions. E.g., it is possible that you toss a fair coin that lands "heads" 2×10^{400} times in a row. Yet it actually seems about as true as anything actually seems true that were you to toss the coin 2×10^{400} times in a row, the coin would not land heads every time. To avoid any misunderstanding, I am not giving reasons for thinking that a counterfactual having a complete antecedent is true only if the antecedent renders the consequent is highly likely.

[24] I've borrowed this argument from Edward Wierenga (1989: 140-8).

(13)   (C1 ◊→ Truth) & (C1 ◊→ Lie).

Propositions (12) and (13) are not clearly equivalent. Libertarians are not obliged to translate "Alice *might* tell the truth in the circumstances" as "(C1 ◊→ Truth)." Rather, they are permitted to translate "Alice *might* tell the truth in the circumstances" as "◊(C1 & Truth)." Thus, instead of warranting an inference to (13), (12) may only warrant an inference to

(12*)   ◊(C1 & Truth) & ◊(C1 & Lie).

However, (12*) does not imply

(14)   ~(C1 □→ Lie) & ~(C1 □→ Truth),

which the success of van Inwagen's argument requires. Furthermore, recall that the only support van Inwagen adduces for (12) is

(11)   Alice (freely) tells the truth in set C1 of complete nondetermining circumstances.

Now the minimal assumption that all libertarians accept—an assumption that they think follows from (11)—is

(Crux)   Necessarily, if an agent Δ freely performs some action A, then the set C of complete circumstances in which Δ acts does not include Δ*'s (freely) performing A*.

Applying Crux to Alice's situation expressed in (12) results only in (12*), which lends credence to the contention that (12*) correctly translates (12). But (12*) neither implies (13) nor implies (17). Thus, there are reasons for thinking that (12) does not imply (13).
Coupling this result with my cumulative critique suggests that the Jury has no reason to accept van Inwagen's Might-Argument for (9). In addition, the Jury should compare their degree of confidence that (12) implies (13) with their degree of confidence in (8). The plausibility of (8) clearly outweighs the plausibility of (12)'s implying (13). Mighty the Might-Argument is not.
Since (9) is presumably a premise in van Inwagen's argument for the impossibility of free will, there are two claims to compare with respect to their levels of intuitive strength. First, the cumulative critique fails, and (12) implies (13) even though there are alternative reasonable ways to interpret (12) that do not imply (13). Second, (8) is true. The first claim, if true, is not as intuitively true as the second claim. By standard, the Jury should not accept van Inwagen's MMCA.
I conclude this section with an interesting consequence that follows from van Inwagen's contention that the move from (12) to (13) is valid. The consequence is this: Worlds in which Alice always makes the same decision (i.e., a steady-world) are just as similar to α as any world (such as the flip-flop-world) where Alice makes different decisions. That is, it is false that if God *were* to replay Alice's situation a very large number of times, then it *would* be the case that *sometimes* Alice tells the truth and *sometimes* Alice lies. For, events prior to the replays do not determine that Alice sometimes tells the truth and sometimes lies. By van Inwagen's lights, it follows that

84

(19)   (Multiple Replays ◊→ Sometimes Truth and sometimes Lie) & (Multiple Replays ◊→
        ~(Sometimes Truth and sometimes Lie)),

which is equivalent to

(20)   ~(Multiple Replays □→ ~(Sometimes Truth and sometimes Lie)) & ~(Multiple Replays
        □→ (Sometimes Truth and sometimes Lie)),

which is equivalent to

(21)   ~(Multiple Replays □→ Same decision every time) & ~(Multiple Replays □→
        Sometimes Truth and sometimes Lie),

which entails that

(22)   With respect to their degree of closeness to α, steady-worlds *tie* with any other world in
        which God replays Alice's situation.

Given this result, one may wonder why some non-steady-worlds (e.g., flip-flop-worlds)
are any more relevant to whether Alice is free in α than a steady-world. Since there is a vast
difference between what goes on in a steady-world and what goes on in a flip-flop-world, we err
in thinking that a non-steady-world represents an arbitrarily chosen sufficiently close world to
α—as vastly different worlds tie in sufficient proximity. A flip-flop-world may be a nearby
world, but so too is a steady-world. It stands to reason that if we draw any conclusions from
entertaining what goes on in a non-steady-world (e.g., a flip-flop-world), then we should be able
to draw the same conclusions from entertaining what goes on in a steady-world. But whether we
can draw the same conclusions from entertaining what goes on in a steady-world is far from
clear. Indeed, the next section argues that we should not draw the same conclusions.

4.5   Third Objection: On What We Could Learn
        So far I've only shown that (9) should be quite controversial from the perspective of the
Jury. Let's pretend that van Inwagen responds by reasoning as follows:

> I cannot get myself to believe that there are true CUEs. But for sake of argument (and
> only for sake of argument), I grant you that there are true CUEs. In particular, I grant
> that there are true counterfactuals of undetermined actions (CUAs). But this is
> irrelevant. Just focus on the flip-flop-world, which you have granted is possible. I
> firmly stand by my claim that, once we witness Alice's flipping and flopping
> throughout the 726 replays, there is nothing we could learn that should undermine our
> conviction that the outcome of the 727th replay will be due simply to chance. That
> there are true CUAs does nothing to undermine the inescapable impression we have
> that in the next replay the action performed is due simply to chance, for the truth of
> each of these CUAs is due simply to chance. And since we are convinced that Alice's
> actions are chancy in the flip-flop-world, then since there's no relevant difference in the
> circumstances in which she acts, we should be convinced that Alice's telling the truth in

α is freedom-squashingly chancy. This conclusion is perfectly general, since, necessarily, for any (putatively free) undetermined action, there is a corresponding flip-flop-world. Thus, we should be convinced that there are no undetermined free actions in any world, which is just to say that we should be convinced that an undetermined free action is impossible.

I do not share the belief that as the number of replays increases so too would our degree of confidence that the outcomes will be due simply to chance. But I guess it is broadly logically possible that I (or any member of the Jury, for that matter) reasonably harbor this conviction. So, suppose that after witnessing 726 replays together, we were convinced that what Alice does in the next replay is simply a matter of chance and that therefore Alice is not *able* to tell the truth and Alice is not *able* to lie.

Nonetheless, there is something that we could learn that may reasonably undermine our conviction, and it has nothing to do with CUEs. We could learn something in the wake of reflecting on our situation thus: We know that the outcome of each replay is undetermined. So, let's think of a nearby world (a nearby world with respect to our flip-flop-world) where we witness Alice *always* telling the truth. In this steady-world, we would have an inescapable impression that the next outcome will *not* be a mere matter of chance. However, since Alice is for all practical purposes in the same set of complete circumstances for each replay, we should only conclude that what she does next is *not* due simply to chance. We should therefore shed our conviction in our own flip-flop-world that what happens next is due simply to chance.

It seems, then, that the strategy of considering God's replaying the universe cuts both ways. To be sure, though, let's unpack this argument in more detail. Imagine that we are suitably placed, metaphysically speaking, to observe the unfolding of a ludicrously large number of replays where Alice performs the undetermined action of telling the truth each and every time. That is, imagine that we witness the unfolding of events in a steady-world. If a flip-flop-world is possible, then a steady-world is possible. Moreover, the validity of the inference from (12) to (13) strictly implies (22). Hence, for any undetermined action, a steady-world is just as close to the flip-flop-world as the flip-flop-world is close to itself. So, after 100 replays, Alice has told the truth 101 times. Is it not true that as we watch the number of replays increase in this steady-world, we shall become convinced that what will happen in the next replay is *not* a mere matter of chance? If we have watched 726 replays, we shall be faced with the inescapable impression that what happens in the 727th replay will *not* be due simply to chance. There is no obvious reason that we should resist this temptation. These assumptions are consistent with van Inwagen's (2000: 15) claim, "Each time God places the universe in this state, both "truth" and "lie" are consistent with the universe's being in this state and the laws of nature." But then, since we know that each set of complete circumstances in which Alice performs an action is *nondetermining*, we reason as follows:

> A sheaf of possible futures (possible in the sense of being consistent with the laws) leads "away" from this state, and, if the sheaf is assigned a measure of 1, <u>surely, we must assign a measure of 1 to the largest sub-sheaf in all of whose members Alice tells the truth</u>. We must make this assignment because it is the only reasonable explanation of the observed approximate dominance of the "truth" over "lie" outcomes in the series of replays. And if we accept this general conclusion, what other conclusion can we accept about the 727th replay (which is about to

commence) than this: each of the two possible outcomes of this replay has an objective, "ground-floor" probability of 1—and there's nothing more to be said? And this, surely, means that, in the strictest sense imaginable, the outcome of the replay will not be a matter of chance.

Now, obviously, what holds for the 727th replay holds for all of them, including the one that wasn't strictly a *re*play, the initial sequence of events. But this result concerning the "initial replay," the "play," so to speak, should hold whether or not God bothers to produce any replays. And if He does not—well, that's just the actual situation. Therefore, an undetermined act is not simply a matter of chance: if it was undetermined in the one, actual case whether Alice lied or told the truth, it was not a mere matter of chance whether she lied or told the truth. So, since we know that each action Alice performed was not determined by antecedent events (i.e., there was no sufficient causal condition for her action of telling the truth, and there was no sufficient causal condition for her action of telling the lie), we acquire the conviction that Alice was both *able* to tell the truth and *able* to lie. For, on each occasion, she determined the outcome of an antecedently undetermined process whose outcome was not a matter of objective, ground-floor chance. And this conclusion is perfectly general, since, necessarily, for any (putatively free) undetermined action, there is a corresponding steady-world. Thus, we should be convinced that there are undetermined free actions in some worlds, which is just to say that we should be convinced that an undetermined free action is possible.

Of course, our argument breaks down at some step. But that is precisely the point. Since our flawed argument straightforwardly mirrors van Inwagen's (2000: 15) "plausible, intuitive" argument, by parity of reasoning, van Inwagen's argument breaks down too.

Consider van Inwagen's (2000: 17) pivotal claim, "Nothing we could possibly learn, nothing God knows, it would seem, should lead us to distrust our initial inclination to say that the outcome of the next replay [in the flip-flop-world] will be a matter of chance." On the contrary, there is something we could possibly learn. More simply, there is something we could learn. In particular, while observing some flip-flop-world we could entertain the argument just advanced above. We learn that the grounds for our initial inclination consist of sand, since we consider an analogous argument for the opposite inclination and reasonably reject them both. I conclude that van Inwagen's MMCA need not oblige the reasonable libertarian to give up her view.

4.6   A Direct Case Against the Mere Matter of Chance Argument

Are there good reasons for thinking that a free undetermined action is possible? Some libertarians think so. Consider the limiting case of God's creating our possible world rather than some other possible world. The objective probability of any particular world being actual is, well, very low. However, the theist already believes that God freely created (or creates) our world. Moreover, God was not determined to create our world. But, God's freely creating our world is possible—*a fortiori* some free undetermined action is possible.

Several philosophers have made this less than maximally appreciated point, philosophers including Thomas Reid, Alvin Plantinga, and Thomas Flint. For example, Reid (1895: 324)

says, "And, if the Deity acts freely, every argument brought to prove that freedom of action is impossible, must fall to the ground."[25] Later, he (1895: 338) continues:

> …[E]very argument [attempting to show the impossibility of free will], if it were really conclusive, must extend to the Deity, as well as to all created beings; and necessary existence, which has always been considered as the prerogative of the Supreme Being, must belong equally to every creature and to every event, even the most trifling. This I take to be the system of Spinosa, and of those among the ancients, who carried fatality to the highest pitch.

The idea, here, is that the Christian reasonably believes that God's act of creating creatures is free. Moreover, if God is determined to create as he does, then, since he exists necessarily, so too everything he creates exists necessarily. However, surely either there are contingently existing objects or might have been other objects that don't in fact exist. Thus, God is not determined to create as he does. So, the Christian believes that there is some undetermined free action.

Plantinga and Flint's observations on this point are equally insightful. Notice that if free will is impossible, then free will (vacuously) requires determinism. On whether free will requires determinism, Plantinga (1990: 31) is worth quoting *in extenso*:

> But the Christian thinker will find this claim monumentally implausible. Presumably the [critic] means to hold that what he says characterizes actions generally, not just those of human beings. He will hold that it is a *necessary* truth that if an agent isn't [deterministically] caused to perform an action then it is a mere matter of chance that the agent in question performs the action in question. From a Christian perspective, however, this is wholly incredible. For God performs actions, and performs free actions; and surely it is not the case that there are causal laws and antecedent conditions outside his control that determine what he does. On the contrary: God is the author of the causal laws that do in fact obtain; indeed, perhaps the best way to think of these causal laws is as records of the ways in which God ordinarily treats the beings he has created. But of course it is not simply a matter of *chance* that God does what he does—creates and upholds the world, let's say, and offers redemption and renewal to his children. So a Christian philosopher has an extremely good reason for rejecting this premise…

Likewise, Flint (1998a: 30) claims:

> God is a free creator. Yet it seems that the typical…complaints against the libertarian notion of a free action are (from an orthodox Christian's perspective) not applicable to God's actions. But then, if God's actions can be rational and

---

[25] Elsewhere (1895: 284), Reid notes, "…[F]or it is a contradiction to say, that a perfect being does what is wrong or unreasonable. But to say, that he does not act freely, because he always does what is best, is to say, that the proper use of liberty [i.e., free will] destroys liberty, and that liberty consists only in its abuse."

appropriate, actions for which he is properly seen as morally praiseworthy, even in the absence of any ultimate causes beyond his control, then there clearly can be no *conceptual* problem with the notion of free, rational, responsible, but undetermined actions. And if there is no such conceptual problem, then there seems to be no conceptual problem with viewing ourselves as agents [who freely perform undetermined actions].

Clearly, if traditional theism[26] is right, at least one obviously free act has no sufficient causal condition. Our ruminations about God's freedom do not *obviously* entail that a free action could not have a sufficient causal condition. For, one may think that God's free activity is undetermined while maintaining that some creature's free action is determined. However, Chapter 3's conclusions imply that there cannot be any free action having a sufficient causal condition.

4.7   Conclusion
        This Chapter discussed van Inwagen's Mere Matter of Chance Argument, the conclusion of which is that an undetermined free action is impossible. Though van Inwagen ultimately believes that his argument should fail, he claims not to have a clue as to how this could be. I attempted to provide several of the missing clues, any one of which is "clue enough." I argued that the Jury (i.e., intelligent, reflective, and reasonable folk who have yet to rule on the success or failure of van Inwagen's argument) should not find van Inwagen's argument for the impossibility of an undetermined free action convincing.
        I argued that van Inwagen's thought-experiment may be incoherent, since perhaps the agent cannot survive God's replaying the course of events, and the laws of nature might not happen to permit God's replaying the course of events. I discussed in detail van Inwagen's rhetorically powerful claim, *viz.*, (9). I argued that there are good reasons for (9)'s falsity. I argued that van Inwagen's argument for (9) fails. I argued that even if (9) is true, van Inwagen's Mere Matter of Chance Argument fails, as there is an analogous argument for a contrary conclusion. Finally, I noted that theists have an overwhelmingly good reason to hold that there actually are undetermined free actions. Hence, some reasonable Conventionals and Flexibles have grounds for thinking that an undetermined free action is possible. If my arguments in this Chapter are cogent, then there is one less reason for thinking that free will is a mystery.
        This is by no means the final word on the status of undetermined free actions. Defending the claim that there might be an undetermined free action counts only as the first step in outlining an adequate account of the phenomena. Moreover, we pre-reflectively think that we are free. But *how* could an undetermined action be up to any *finite* creatures like us who are limited in power while being causally influenced by our circumstances? At the very least, an adequate account of free will should account for how we perform paradigmatically free actions that are nonetheless undetermined by the circumstances that causally contribute to our acting as we do. In the absence of such an account, perhaps the most we can warrantedly believe is that even though there are free undetermined actions, it is a mystery (to us) how this could be. In the balance of this book, I try to diminish this mystery as much as I can.

---

[26] See Clive Staples Lewis (1943).

CHAPTER 5

THREE LIBERTARIAN POSITIONS

I change my mind about the problem of free will every time I think about it, and therefore cannot offer any view with even moderate confidence; but my present opinion is that nothing that might be a solution has yet been described. This is not a case where there are several possible candidate solutions and we don't know which is correct. It is a case where nothing believable has (to my knowledge) been proposed by anyone in the extensive public discussion of the subject.
—Thomas Nagel (1995: 35)

Chapters 3 and 4 confirm the assumption that incompatibilism is true and that free action is possible. Indeed, from here on out, I assume that some finite agent really does perform a free action. So, we assume libertarianism. We turn, now, to the Intelligibility Question. How is one supposed to understand how it is that an agent performs an undetermined free action? Is an undetermined free action coherent and intelligible?

This chapter sketches the three major libertarian candidates for answering the Intelligibility Question: simple indeterminism, causal indeterminism, and agent-causation. I reason that the former two views have trouble meeting the origination condition, a condition required by an agent's directly free action. An agent's *directly* free action just is a free action the freedom of which is not derived from the freedom of any other action the agent performs. My strategy, then, is to argue that simple indeterminism and causal indeterminism do not account for an agent's being an underived source of any part of her directly free action. To this end, I review[1] what others say about this condition of origination, and I attempt to put a sharper point on their intuitions and convictions by formulating the condition of origination with more precision.

5.1   The Condition of Origination: Whenceness

Philosophers typically argue for or against certain accounts of free action by focusing on the notion of control or on an agent's power to do otherwise. This is unfortunate. We do not have a clear enough understanding of 'control' for the term to do duty in weighty philosophical argumentation. Since the term 'control' may be used in so many different ways, I find it extremely difficult to see how using the term advances the discussion regarding the nature of free action. Thermostats control a region's temperature. Wires control the flow of electricity. There are control consoles for launching rockets. People use the term 'control freaks', and this is perhaps not to highlight one's freedom but lack thereof. One controls which direction one's car

---

[1] See Chapter 2, §2.2.2.

travels by moving the steering wheel, but it's not at all clear that such control has anything to do with one's acting freely. Other examples are fairly easy to imagine.

In contrast to the notion of control, I think we do have a firm enough grasp of an agent's power to do otherwise, a phenomenon expressed by what I have been calling the condition of dual power or dual ability.[2] Carl Ginet (1995: 69; 1990: 9) sums up the condition of dual power, saying, "By a *free* action I mean one such that until the time of its occurrence the agent had it in her power to perform some alternative action (or to be inactive) instead."

Unfortunately, while nearly all experts on free will focus almost exclusively on the condition of dual power, they overlook or fail to appreciate yet another intuitive feature of free action. The idea is that when acting freely, an agent is an ultimate source, underived originator, or initiator of something. In acting freely, she is a self-determined, self-directed, sole author of change in the world. And even when experts of free will mention ultimate origination, they hardly put the concept to work in rigorous philosophical argumentation.

Free action intuitively requires origination. As we saw in Chapter 2, §2.2.2, this pre-theoretic feature of free action can be found throughout the literature. But virtually no one is making use of it when constructing arguments,[3] and no one to my knowledge has formulated the origination condition with much precision. Since I shall use the notion of origination/self-determination in arguing that agent-causation outshines simple indeterminism and causal indeterminism, we shall need an intuitive formulation of the condition of origination/self-determination. Permit me first to review what several philosophers pre-theoretically have to say about origination:

- Gary Watson (1987: 145). "Any reasonable conception" of free will must capture the right interpretation "of two different features of freedom…—namely, self-determination (or autonomy) and the availability of alternative possibilities. Even classical compatibilism made room, or tried to make room, for both."

- John Locke (1894: Book 2, Chapter 21, §48). If someone is "under the determination of some other than himself" then there is "want of liberty."

- Benedict Spinoza (1677: Point 1, Proposition 11). "That thing is called 'free' ... is determined to act by itself alone."

- Peter van Inwagen (1983: 11, *his emphasis*). "[T]he concept of an agent's power to act [freely] would seem not to be the concept of a power that is dispositional or reactive, but rather the concept of a power to *originate* changes in the environment."

- Daniel Dennett (1995: 55 note 1). "How can a person be an author of decisions, and not merely the locus of causal summation for external influences?"[4]

---

[2] See Chapter 2, § 2.2.1.

[3] Exceptions may be Randolph Clarke (2003) and Kane (1996a).

[4] Dennett takes this question to be among an incomplete list of questions that composes the problem of free will.

- Robert Kane (1996a: 4, 15, 78, 196). Free will is "the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes."

- Randolph Clarke (1996a: 26). "An agent's having multiple courses of action that she *might* perform does not suffice for the agent's having freedom; rather, the agent must exercise a causal power in such a way that she is a source of her activity."

- Ted Honderich (1993: 2). "Each of us is supposed to have a kind of personal power to originate choices and decisions."

- Honderich (1988: 389). We have a pre-philosophical and pre-theoretical "idea or whatever of a determinate centre, a self, which is uncaused in its activity and which is not superfluous."

- Gary Watson (1987: 146). An agent's free action requires that her action be self-determined in the sense that she herself makes the difference as to whether or not she performs that specific action.

- Clarke (2003: 15). When an agent acts with free will she "…determines, herself, what she does; she is an ultimate source or origin or initiator of her behavior."

- Timothy O'Connor (1996: 145). In acting freely, "…I am quite literally the cause (source, point of origination) of my own behavior."

- O'Connor (1996: 143, 145-6). We are "self-determining" beings. When we freely make choices for consciously considered reasons, how we act is "up to us."

- Laura Ekstrom (2000: 3). A free agent is "fully *self-directed* in her action."

- Thomas Reid (1895: 602). "If the person was the cause of that determination of his own will, he was free in that action, and it is justly imputed to him, whether it be good or bad."

- Michael Zimmerman (1988: 24-5). "If one strictly freely wills an event, then (and only then) one's volition is, in a sense, 'truly one's own'; one is, in a sense, the 'source' of one's volition (and hence of the action that comprises it)."

- Robert Nozick (1981: 313). "We want it to be true that in that very same situation we could have done (*significantly*) otherwise, *so that* our actions will have *originative* value."

- Susan Wolf (1990: 10). "[T]here is a requirement that the agent's control be ultimate—her will must be determined by her self, and her self must not, in turn, be determined by anything external to itself."

- C.A. Campbell (1957: 156-7, 160; 1966: 131-5; and 1967: 48-9). A free agent is "the *sole* author" and that "the *self* determines" her action.

- Thomas Flint (1998a: 23). "Self-determination lies at the very heart of freedom; to say that an act of mine was free, but was ultimately determined by someone or something other than me, someone or something whose determining activity was utterly beyond my power to control, is to speak nonsense."

- Alvin Plantinga (1990: 31). "[T]he notion of a person as an ultimate source of action" lies at the heart of certain important controversies regarding free will.

I wish to make two points in light of these quotations. First, it is important to realize that none of these philosophers argue for these claims. The citations above are not conclusions. Rather, these thinkers register their intuitive convictions regarding the pre-theoretic nature of free action. They deploy their claims as pre-philosophical data. Until these common-sense beliefs are shown mistaken and shown mistaken from beliefs even more certain, we may safely assume that these philosophers have latched onto something quite significant. Clearly, there is a deep-seated intuition that an agent's free action strictly implies that she is self-determining, self-directed, and originative whilst performing the action. Free agents are fundative entities.

Second, while there may very well be an important and informative relation between the condition of dual power and the condition of origination, the connection is not obvious. In most of the texts from which the above quotations were taken, there is not any clear and necessary connection between these two conditions. For example, Watson (1987: 145) lists them as distinct, and Clarke's (1996a: 26) remark suggests that the condition of dual power does not imply origination. At the very least, Clarke suggests that an agent may perform one of several causally unclosed actions yet not be an originator/source of her action. Moreover, most of the quotations above make no mention of what I am calling dual power.

Given these various pre-philosophical intuitions, I propose the following proposition, which I think expresses the condition of origination that free action requires:

(Whenceness)   Necessarily, for any agent Δ performing any directly free action A, Δ is an underived/ultimate originator of an essential element/part of A.

Whenceness states a necessary condition for an agent's directly free action. Clearly, every finite agent performs a free action only if the agent performs a *directly* free action.

Notice that for any action A, A is an essential element/part of A—not a proper part, but a part nonetheless. Hence, if an agent is an underived originator of action A, then she is an underived originator of an essential element of A. Whenceness, then, is neutral with respect to whether an agent is an originator of a *proper* element/part of A, on the one hand, or whether the agent is an originator of the entire act A, on the other hand.

That Whenceness captures the intuitions canvassed above seems clear. For example, if agent Δ is an underived originator of an essential part of action A, then clearly Δ is self-determining in virtue of the fact that either Δ determines herself that A occur by originating A or else A strictly implies that Δ determine that an essential and proper part of A occurs. Moreover, if *Δ's performing A* is determined by something else, then clearly Δ will not be an *underived/ultimate* originator of an essential component of A. This confirms Chapter 3's contention that libertarianism is possible. Therefore, meeting Whenceness implies that an agent is self-determining whilst performing her directly free action. Whenceness brings into relief the

intuitions regarding the notion of self-determination advanced by Watson, Locke, Spinoza, Clarke, Wolf, Reid, Ekstrom, O'Connor, Campbell, and Flint.

Satisfying Whenceness also clearly confirms van Inwagen's intuition that a free agent's power is a power to *originate* change in the world. If *Δ's performing some directly free action A* essentially implies that Δ is an underived originator of an essential element/part of A, then obviously she exercises a power to originate change in the world. Likewise, Δ would be an ultimate source, author, and creator. Therefore, Whenceness confirms the intuitions of Dennett, Kane, Clarke, Zimmerman, Plantinga, O'Connor, and Honderich.

Meeting Whenceness, then, would be a fine achievement for a theory of free action! If we learn that it cannot be done, then we should abandon such a high standard for adequate theories of free action. But if one theory provides a unique and principled way to meet Whenceness, and if this theory has no special and intractable problems, then we have a good reason for adopting this theory.

5.2   Simple Indeterminism

Simple indeterminism (SI) is a rather minimalist libertarian view, positing no positive causal requirements on one's free action. The only causal requirement is negative, stating that every free action *not* be causally determined.[5] The chief advocates of SI are Carl Ginet (1997, 1995, 1990), Stewart Goetz (2000, 1997, 1988), and Hugh McCann (1998). I criticize the accounts of Ginet and Goetz.

5.2.1   Carl Ginet

Ginet (1997, 1990) characterizes his view by giving a theory of action and then simply adding that a *free* action is an action that just happens not to be causally determined. Hence, a free action is simply an action not included by the complete circumstances in which it is performed.

Call a *basic* action an action the performance of which does not require performing any other action. Overt actions, such as lifting a wineglass or taking out the trash, require performing other actions. For example, lifting a wine glass may include other another action such as choosing to lift the glass, intending to lift the glass for the purpose of acquiring a buzz, reaching for the glass, and the like. Recognizing this intuitive yet hard to specify distinction between basic and non-basic action, Ginet thinks that every action either is or causally begins with—in the right sort of way—a causally simple mental action. A causally *simple* mental event has no internal causal structure. A causally *complex* event, in contrast, is *something's causing an event*.

For Ginet, a simple mental event is an action precisely when it has a certain type of intrinsic phenomenal quality, which he calls the "actish phenomenal quality." Ginet (1997: 89) explains:

> [T]he simple mental event of my volition to exert force with a part of my body phenomenally seems to me to be intrinsically an event that does not just happen to me, that does not occur unbidden, but it is, rather, as if I make it occur, as if I

---

determine that it will happen just when and as it does (likewise for simple mental acts that are not volitions, such as my mentally saying "Shucks!").

According to Ginet, so long as a simple mental event *seems* to one *as if* one is producing it or bringing it about or making it occur, then the mental event is one's action. The act of lifting a wineglass has at its root a mentally simple action, such as a volition to lift the glass.[6] This volition then causes, in a nondeviant way, other events that, taken as a whole, constitute the action. These simple mental actions are the only candidates for being directly free actions.[7]

Ginet maintains that free actions are sorted from unfree actions in virtue of whether or not the complete circumstances in which they are performed include the actions. If the circumstances include the action, which means that the action is causally determined, then the action is not free. If the circumstances just happen not to include the action, then the action is free. Ginet (1997: 89) notes,

> A simple mental event's having this intrinsic actish phenomenal quality is sufficient for its being an *action*. But its having the quality entails nothing either way as to whether it satisfies the incompatibilist requirement for *free* action (which is that it not be causally necessitated by antecedent events).

One might wonder how an agent is *self*-determining in performing her directly free action, given that her action is not causally determined by anything whatsoever.[8] The question, here, concerns how Ginet's theory is supposed to meet Whenceness. According to Ginet's account, how is an agent supposed to be an ultimate originator of an essential part of her directly free action? There is a hint of an answer when Ginet (1997: 85) characterizes

---

[6] For McCann (1998: 163), basic actions are intrinsically intentional. For example, when one decides, one means to decide as one does decide.

[7] The thorny problem of "deviant" or "wayward" causal chains is important but irrelevant to our study. The problem is that an event might not be an action even though it is caused by the right kinds of mental events. Richard Taylor (1992: 249) portrays the problem thus:

> Suppose, for example, that a member of an audience keenly desires to attract the speaker's attention but, being shy, only fidgets uncomfortably in his seat and blushes. We may suppose, further, that he does attract the speaker's attention by his very fidgeting; but he did not fidget in order to catch the speaker's attention, even though he desired that result and might well have realized that such behavior was going to produce it.

Finding a general solution need not concern us because our focus is only on whether the simple indeterminist can account for directly free actions, which in this case are certain initial mental events that are not caused by previous actions, intentions, or the like. I assume there is a solution to the problem of wayward causation. Advanced, intricate, but arguably inadequate solutions are discussed in John Bishop (1989: Chapters 4 and 5), Myles Brand (1984: 17-30), Donald Davidson (1973: 153-4), Ginet (1990), and Mele (2000, 1992).

[8] Even Ginet (1997: 98) admits that buying into agent-causation would allow us to take the actish phenomenal quality seriously. It would allow us to take this its-seeming-as-if-I-directly-cause-it as a literal perception of reality; it is this that makes so pre-reflectively appealing the idea that what makes a person the agent of an actish event is that she directly causes it.

95

simple indeterminism [as] the view that the occurrence of a simple mental event is determined by its subject if it possesses the 'actish' phenomenal quality and is undetermined by antecedent events.

Timothy O'Connor (2000a: 25) voices the following complaint against Ginet's SI:

The fact that free actions have uncaused volitions at their core is prima facie puzzling. If it is uncaused, if it is in no sense determined to occur by anything at all, then it is not determined to occur by me in particular. And if I don't determine it, then it's not under my control.

Again, I am not clear about what 'under my control' means, but we may simply supplant O'Connor's last proposition with 'And if I don't determine the volition or the volition is not inherently my determining something, then I am not self-determining.'

Ginet rejects O'Connor's enthymematic premise that an uncaused action is in no sense determined to occur by anything at all. For, he thinks that certain uncaused actions are determined by the agent in virtue of simply being the agent's actions. Ginet (1997: 87) explains:

But if [event] *e is* her own free action, then she makes it the case [i.e., she determines] that *e* occurs, not by causing it, but by simply performing it (This latter "by" is logical rather than causal; we have a causal "by" in "I made a C major chord sound by pressing those three keys;" we have a logical "by" in "I made a C major chord sound by making sound simultaneously a C, an E, and a G").

Ginet (1997: 89) continues:

I make my own free, simple mental acts occur, not by causing them, but simply by being their subject, by their being my acts. They are *ipso facto* determined or controlled by me, provided they are free, that is, not determined by something else, not causally necessitated by antecedent states and events.

I find Ginet's reply inadequate. I contend that Ginet's theory fails to meet Whenceness. Consider, for reductio, a case where Ginet's theory holds for a finite agent Δ's directly free action A. Act A is, on his view, a simple mental event—e.g., a volition performed with the aim to satisfy some further goal. Since A is an action, Ginet thinks that A has peculiar actish phenomenal quality. Δ feels as if she produces A. Moreover, since Δ freely performs A, A is not causally necessitated by antecedent events. That is, A is causally *un*determined. Is this enough to satisfy

(Whenceness) Necessarily, for any agent Δ performing any directly free action A, Δ is an underived/ultimate originator of an essential element/part of A?

I think not.

Add the following details to Ginet's conditions. Agent Δ is a morally respectable individual. Unfortunately, an evil demon, Screwtape, maliciously wants Δ to perform some

morally detestable action A. Screwtape has no power over the laws of nature and knows that these laws preclude any sufficient causal condition for *Δ's performing A*. So, Screwtape is unable to arrange things such that *Δ's performing A* is causally determined. But from Screwtape's perspective, this fact is inconsequential. Screwtape still forces Δ to have a certain state of mind (M) that greatly conflicts with Δ's normal psychological makeup. Screwtape knows that the laws of nature imply that *Δ's having M* makes it 99.99% likely that Δ performs A. Due to the quantum indeterminacy of some subatomic particle in Δ's brain, there is a 0.01% chance that Screwtape's plan fails. But Screwtape knows this. And he knows that M will in fact indeterministically cause A. In short, Screwtape arranges Δ's circumstances so that Δ's externally infused M causally produces A.[9] Finally, Screwtape forces Δ to feel as if she is producing A, giving A that actish phenomenal quality.

Now for the question: in this situation, is Δ obviously an ultimate originator of an essential part of her action A? Is Δ clearly an ultimate source of an essential part of A? No. If anything is clear, Δ is not an ultimate originator of an essential part of A. If anything originates either A or any part of A, then either Screwtape does or Δ's conflicted and uncharacteristic mental state M does. Given Δ's circumstances, we should be radically skeptical about Δ's being able to save herself from performing A. Yes, A is an action and has an actish phenomenal feel. A is also causally unclosed by Δ's circumstances. Even so, while performing A, Δ is not obviously an ultimate source of change in the world. Δ is not self-determining in performing A. Therefore, Ginet's conditions fail to satisfy Whenceness. Thus, since Whenceness states a necessary condition on free action, Ginet's theory fails as an adequate account of free action.

Notice that Ginet should not hold that *Δ's determining A* is an essential feature of action A, for any agent-action pair Δ and A. For, suppose either that there is a sufficient causal condition of *Δ's performing A*, or that God directly causes *Δ's performing A*. Again, Ginet thinks that simple mental actions might be determined. Either way, Δ would still be the subject of A, and A would be Δ's action. We stipulate that A, though determined either by a prior sufficient causal condition or by God, still possesses all of its original intrinsic features. So, the simple mental action A still feels to Δ as if she produces it. Action A still has that actish phenomenal quality from Δ's point of view. Nevertheless, while something determines A, it is not Δ. Rather, God or the deterministic cause of A determines A. Moreover, if Δ does not determine A, then simply removing the determining factor (from a conceptual point of view) should not thereby ensure that Δ automatically is self-determining. At the very most, Δ would be not-determined-by-another. But being-not-determined-by-another need not guarantee being-self-determined. And since one's being an ultimate/underived originator of an event entails that one self-determines this event, Ginet's account does not guarantee that an undetermined simple mental event meets Whenceness.

To see this point from another angle, compare two agents, Δ* and Δ, in qualitatively indistinguishable complete circumstances. That is, imagine that Δ* performs simple mental action A* in circumstances C*, and Δ performs simple mental action A in complete circumstances C. Suppose that C* includes *Δ*'s performing A*. For example, C* or (at least a part of C*) deterministically causes A*. In contrast, suppose that C does not include *Δ's*

---

[9] Ginet (2002: 397) thinks that a free mental act need not have any cause. But it is consistent with his view that a free act be caused provided that the causation is not deterministic. See also Ginet (2002: 403, 1997: 92).

performing A. The laws of nature in this C-world permit both C to obtain while *Δ's performing A* fails to occur.[10]

Now, if Δ* is not self-determining in performing A*, how could Δ be self-determining in performing A? If Δ* is not an ultimate originator in performing A*, how does a mere absence of antecedent determination imply that Δ is an ultimate originator in performing A? If we compare these two agents, we find no principled reason to bestow upon Δ the property *being self-determining*.

Ginet's theory, then, raises thorny questions. In virtue of what does Δ determine her simple mental action? How is Δ *self*-determining in performing a simple mental action? The answer cannot be that Δ is the action's subject. For, Δ may be antecedently determined to be the subject of an action, and being deterministically caused to act intuitively precludes being self-determining. So, an account of action *per se* cannot capture how Δ is an ultimate originator. At the most, Ginet's theory implies that an agent is not-determined-by-another.

We're seeking a principled account of ultimate origination. Ultimate origination strictly implies self-determination. Though Ginet's theory implies anti-determination-by-another, it does not obviously imply self-determination. Thus, Ginet's theory does not obviously imply ultimate origination. Ginet's theory, then, simply provides too thin an account of origination for an adequate theory of free will.

5.2.2   Stewart Goetz

Although Stewart Goetz's SI is more sophisticated than Ginet's view, I argue that Goetz also fails to provide a principled theory that meets Whenceness. For purposes of the present discussion, the important features of Goetz's (1997) account run as follows. Necessarily, any agent Δ's directly free action A is Δ's choice, a causally simple mental event. Goetz (1997: 196) asserts, "(i) a choice is the *exercising* by an agent of his mental *power* to choose, where (ii) the exercising of a mental power is essentially an uncaused event." Goetz (1997: 196) stipulates, "A mental power is an ontologically irreducible property which is exhibited by an entity." A mental power is exhibited by an entity precisely when the entity or agent exercises the power. An exercising of a power is acting, and a basic mental action is an exercising of a mental power. When an agent exercises his mental power to choose, he chooses. Goetz (1997: 197) explicitly rejects Ginet's requirement that every mental action has an intrinsic feel to it. Instead, mental acts are simply exercisings of mental powers.

There are other interesting features of Goetz's view—features, e.g., that outline how an agent freely acts for a reason, purpose, or telos. Nonetheless, even setting aside considerations about reasons-explanation, we can already see how Goetz would reply to my criticism of Ginet's theory. Recall Goetz's (i) and (ii):

(i)   □∀x∀Δ{x is Δ's choice ⊃ x is Δ's exercising of his mental power to choose},

(ii)   □∀x{x is an exercising of a mental power ⊃ x is uncaused}.

---

[10] I apply this argument against CI as well. See §5.3.4 and §5.3.5. The argument is inspired by Clarke (2003: 106), though to my knowledge he does not use it against SI.

98

Moreover, Goetz holds that every directly free action is essentially a choice, i.e.,

(iii)   □∀x∀Δ{x is Δ's directly free action ⊃ x is Δ's choice}.

By 'uncaused', Goetz means 'not caused efficiently'. For Goetz (1997: 208 note 8) claims, "In this paper, what is meant by 'causation' is efficient causation which is the exercising of a causal power that produces its effect."
        With only a moment's reflection, we can see that (i) – (iii) implies that any directly free action is essentially uncaused. Thus, our argument against Ginet's theory cannot obviously be translated into a criticism against Goetz's account. For example, Screwtape cannot know that agent Δ's mental state M will indeterministically efficiently cause Δ to perform directly free action A. For, on Goetz's view, A is a choice, and choices cannot be efficiently caused by anything. Also, our comparison between a determined agent's acting and an undetermined agent's acting, where in each case the act is caused, misfires when applied to Goetz's theory. Again, on Goetz's view, a directly free action is a choice and choices do not *just happen* to have no efficient cause; rather, choices essentially have no efficient cause.
        (i) – (iii), then, provide Goetz with considerable mileage. Are they obviously true? They do not seem self-evident, at least to me. In particular, I'll focus on (ii), as it does not strike me as obviously correct. Fortunately, Goetz (1997: 197) offers a rather compressed argument for (ii):

> Support for (ii) is conceptual in nature. An event which is efficiently caused is produced by that cause and as such is an occurrence with respect to which its subject is essentially passive. An event is being made to occur to the subject and it is not active with respect to that event. Since an exercise of mental power is active in nature, it is not produced and, thus, cannot be caused.

So, how is Goetz's argument supposed to work? Here is my attempt to extract his line of thought:

(iv)   □∀x∀y{y efficiently causes x ⊃ y produces x}. [BASIC]

(v)   □∀x∀Δ{((Δ is a subject of x) & ∃y(y produces x)) ⊃ Δ is passive with respect to x}. [BASIC]

(vi)   □∀x∀Δ{Δ is passive with respect to x ⊃ (x is being made to occur to Δ & Δ is not active with respect to x)}. [BASIC]

(vii)   □∀x∀Δ{(x is being made to occur to Δ & Δ is not active with respect to x) ⊃ x is not Δ's action}. [BASIC]

(viii)   Thus, □∀x∀Δ{(Δ is a subject of x) & ∃y(y efficiently causes x)) ⊃ x is not Δ's action}. [From (iv)-(vii)]

(ix)   Equivalently, □∀x∀Δ{(Δ is a subject of x & x is Δ's action) ⊃ ~∃y(y efficiently causes x)}. [From (viii)]

(x)   □∀x∀Δ{x is an exercising of a Δ's mental power ⊃ (x is Δ's action & Δ is the subject of x)}. [BASIC]

(xi)   Thus, □∀x∀Δ{x is an exercising of a Δ's mental power ⊃ ~∃y(y efficiently causes x)}. [From (ix), (x)]

(ii)   Thus, □∀x{x is an exercising of a mental power ⊃ x is uncaused}. [From (xi)]

The argument is obviously valid, having five underived/basic premises. While I believe four of these basic premises, I think (v) is false. (v) captures Goetz's claim that a produced event is an occurrence with respect to which its subject is essentially passive.

Now, (v) is not self-evident and seems quite controversial. After all, suppose that event *e* is directly produced by the substance who is Δ herself. Perhaps *e* is some mental state of Δ, a simple mental event. In this case, it is not at all obvious that Δ is passive with respect to *e*. On the contrary, Δ is active with respect to *e* by virtue of directly producing *e*.

Or suppose that Δ's very own settled psychological constellation (DK) of desires and known beliefs produce *e*. Perhaps even *DK's producing e* counts as an action that Δ performs. If this is possible, then we have a case where Δ is not passive vis-à-vis *e*. On the contrary, Δ would be active with respect to *e*, as *e* would be an essential part of Δ's action.

Consider also Randolph Clarke's (2003: 27 note 28) comment on (v):

> No argument is provided to back up the claim [i.e., to back up (v)], and I for one do not find it self-evident. On the contrary, an intention-acquisition's being caused, in an appropriate way, by the individual's having relevant beliefs and desires and an intention to make up her mind seems to me a good candidate for an exercise by that individual of active control, and such an intention-acquisition seems a good candidate for any event with respect to which that individual is active, precisely because of the way in which (and that by which) it is caused.

I agree with Clarke. Goetz's premise (v) is not self-evident and we have some reason for thinking (v) is false. I conclude that (v) is too tenuous to be a load-bearing premise. Possibly, something could produce an event *e* with respect to which its subject is active, e.g., the subject herself could cause *e*.

Moreover, Goetz's position has monumentally controversial consequences. Recall he (1997: 196-7) asserts that every basic mental action is essentially an exercising of a causal power. So, Goetz's view implies that, necessarily, determinism is true only if there is no mental action. That is, determinism strictly implies that no one ever mentally does anything. For, suppose that an agent acts mentally and determinism is true. So, there is something that produces the agent's mental action. But, on Goetz's view, nothing can produce a mental action. Hence, something does, yet nothing can, produce the agent's mental action, which is absurd. But surely it's at least possible that someone be determined to act mentally. Couldn't one be deterministically caused to think, to deliberate, to entertain propositions, to intend, to consider, or to concentrate? I suspect virtually every compatibilist would think so. If (v) is true, they are all mistaken. In some determined worlds, people look like they're acting, they have the same phenomenology involved when acting, and they believe they are acting. But in fact there are no

100

doings, if Goetz is right. There's no way these determined people could be right. They could not know that they act because determined agents cannot act. If they hold such beliefs, they are *bound* to be wrong.

So, if Goetz is right, then we have an even simpler argument for incompatibilism, assuming that freely acting requires mentally acting. Mentally acting precludes determinism. Hence, freely acting precludes determinism. Is it at least broadly logically possible that something efficiently cause someone to act mentally? It just seems so.

I conclude that Goetz's argument for holding that, necessarily, one's directly free action has no efficient cause fails. Of course, Goetz may simply stipulate that this is part of his view. But I should think one would like a more principled reason to hold to such a controversial thesis. All things considered, maybe adopting Goetz's stipulative way of handling Whenceness is the best we can do. Perhaps another libertarian view fares better than SI, offering a more principled account of free will.

5.3   Causal Indeterminism

We have seen that SI, requiring no positive causal constraints on free action, fails to satisfy Whenceness. We now turn to assess a family of positions falling under the title *causal indeterminism (CI)*. Every member of this family implies that a free action is undetermined by antecedent events but nonetheless is caused by (but only by) certain of these antecedent *events*. The term 'caused', here, is taken to mean 'efficiently caused'. Also, we define the term 'produce' by saying that x *produces* y *iff* x efficiently causes y. So, causal indeterminists claim that an agent's directly free action A *must be produced* by the set C of complete circumstances in which it is performed, but nonetheless C does not include A's occurring. How is this supposed to work for an agent's performing a directly free action?

Causal indeterminists hold that a directly free action A *happens to be produced* by special agent-involving mental events—events such as the agent's having both desires that favor performing A and beliefs to the effect that performing A would help satisfy those desires. Nonetheless, those very reasons (i.e., those very beliefs and desires) might have failed to produce the action they in fact caused.[11] So, the causation of an agent's free action by her having reasons, according to the causal indeterminist, is indeterministic causation rather than deterministic causation. The laws of nature just happen not to guarantee that the agent's action follow her having those reasons. The laws of nature impose a statistical or probabilistic constraints on the agent's reasons' issuing in one action rather than an alternative action.[12]

So, given that the agent's complete circumstances do not strictly imply that the agent performs A, the agent's refraining from performing A is not causally closed, and of course it is not causally closed that the agent perform A, since by hypothesis the agent performs A. Up until

---

[11] The agent's reasons may be any combination of a number of various types of *psychological attitudes*, which need not be limited to beliefs and desires. As Kane (1996a: 28) notes, "[W]e may cite, as her reasons for choice, her wants or desires; preferences, factual beliefs and expectations; likes and dislikes; prior intentions; interests and memories; normative and evaluative beliefs; fears, hopes, and other emotions; and so on." For sake of simplicity and in keeping with the current literature on reasons-explanations, I let *reasons* correspond to some belief-desire complex or part thereof.

[12] I find the notion of indeterministic causation difficult to grasp, but I assume it is coherent while I explain and critically evaluate CI. I discuss the notion of causation in more detail in Chapter 6.

the agent performs A, her future is not causally settled. It is not causally fixed in advance. What she does might not have been done in the very same complete circumstances in which she acted. So she might have done something else, even if it is simply refraining from performing the action she actually performs. Consequently, it seems that the agent has, at least in some minimal sense, the ability to do otherwise. Moreover, that some of the agent's own beliefs about her situation and her own desires that favor her performing A together produce action A strongly suggests that the agent is acting *on*, *for*, or *from* those beliefs and desires. It seems, then, that the agent rationally performs A, as she acts on the basis of her own beliefs and desires (i.e., her own reasons). O'Connor (2000a: 29) summarizes the basic causal indeterminist view, stating:

> It holds that in many circumstances, persons have distinct desire-belief complexes, or reasons, that point toward different courses of action and that the performance of any of these—not just the one chosen—would coherently 'graft' onto precisely the same prior circumstance. Each such reason is a potential cause of the corresponding behavior, and whichever action is undertaken will have been caused by the matching reason(s). In no case does an action occur that is uncaused.

So, on the causal indeterminist's view, an agent's free action A is produced by her having reasons favoring that action. But the agent's having those reasons need not have produced A. The possession of other reasons, which she did in fact possess, might have produced a different action B. In typical (non-Frankfurtian) cases, had the agent performed the alternative action B that was not causally closed to her, then B would have been produced by her having a different belief-desire complex (a different set of reasons), where she had these other reasons even at the time she actually performed A.

Consider the following working example. Juan grew up and still lives in the small village of Huachichil, Mexico. Juan encounters an unprecedented opportunity to migrate to Nogales, Mexico, a border town between Mexico and Arizona. If he migrates to Nogales, he can immediately begin working in a new factory that General Motors moved just across the border. He desires to carve out a future for himself and break the cycle of poverty that has imprisoned his family for centuries, and he believes that migrating to Nogales can best satisfy this desire— call Juan's having *this* desire-belief complex *DB*. On the other hand, , if Juan stays in Huachichil, he's sure to continue to live as a poor farmer. However, he can take care of his aging grandparents and younger sister. They desperately need someone to care for them. He desires to take care of his family and believes he can best satisfy this desire by staying in Huachichil—call Juan's having this desire-belief complex *DB**.

Now suppose that Juan is torn but deliberates very carefully about what to do. He must choose between staying in Huachichil or else migrating to Nogales. We assume Juan's complete circumstances include neither choice/decision. As things happen to turn out, he chooses to migrate to Nogales. His act of choosing to migrate is produced by his desire to gain financial autonomy and his belief that working in the GM factory will contribute to satisfying this desire. That is, DB indeterministically produces his choice to migrate. In keeping with CI, we suppose that had he chosen to stay in Huachichil, *this* choice would have been produced by a combination of his having a desire to take care of his family together with his having a belief that staying in Huachichil best satisfies *this* desire. That is, had he chosen to remain in Huachichil, DB* would

have indeterministically produced his choosing to remain in Huachichil. According to causal indeterminists, since Juan's choosing to migrate is causally undetermined by preceding events but nonetheless is produced by his own reasons favoring the choice he makes, Juan's choosing to migrate to Nogales is directly free.

5.3.1 Robert Kane on Origination

To see how CI is lacking, discussing a particular causal indeterminist view behooves our study. Experts on free action regard Robert Kane's causal indeterminist view as one of the most detailed and sophisticated libertarian accounts.[13] This section argues in detail that Kane's view intuitively possesses insufficient resources to account for an agent's originating any essential part of her free activity, i.e., for satisfying Whenceness. The next section advances the Argument from Origination, which targets causal indeterminist views but may also be seen as posing a general challenge to any libertarian account. We'll see that an agent-causal view has the best resources to meet this challenge.

Kane (1996a: 4) defines free will of agents as their "power to be the ultimate creator (or originators) and sustainers of their own ends or purposes." He believes that experts on free action have focused on the condition of dual ability while unfortunately ignoring what I have dubbed the origination condition (expressed by Whenceness), which he sometimes calls *ultimate creation*, *underived origination*, or *sole authorship*.[14] The terminology here is rhetorically illuminating. Kane apparently attempts to account for how an agent originates something implied by her directly free activity. Indeed, Kane (1996a: 59, 77) suggests that if there is going to be any advance beyond the stalemate between incompatibilists and compatibilists, the former will need to show that the latter cannot meet the origination condition.[15]

Kane (1996a: 35, 72) attempts to capture a necessary condition of being an ultimate creator or originator with his condition of *ultimate responsibility*:

---

[13] "[Kane's (1996a) book] is, quite simply, the best presentation of the 'positive' aspect of libertarianism I have ever read," writes John Martin Fischer (2000: 141, 148), "It will help shape the course of research—especially on libertarianism—for years to come. It is a significant achievement." Randolph Clarke (2003: 30n2) says, "Kane…has set out the most sophisticated version and presented the most sustained defense of a libertarian account of this type."

[14] See Kane (1996a: 4, 33-7, 58-9, 64-78, 79-80).

[15] I do not share this belief.

(UR) An agent is *ultimately responsible* for some (event or state) E's occurring only if

- (R) the agent is personally responsible for E's occurring in a sense which entails that something the agent voluntarily (or willingly[16]) did or omitted, and for which the agent could have voluntarily done otherwise, either was, or causally contributed to, E's occurrence and made a difference to whether or not E occurred; and
- (U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y.[17]

Kane (1996a: 35) says, "The basic idea is that the *ultimate responsibility* lies where the *ultimate cause* is." For, "U requires for ultimate responsibility that the responsibility required by R must 'backtrack' to the sources of the agent's responsible actions." For finite creatures like us, this backtracking stops with special choices that we voluntarily make. Kane (1996a: 75, 125) calls them *self-forming actions* (SFAs).

Since my focus concerns *directly* free actions, which are by definition non-derivatively free, I shall focus on whether SFAs are free. On Kane's view, SFAs just are directly free actions. In particular, I focus on whether Kane's self-forming actions intuitively fulfill Whenceness. Since I focus on self-forming actions, I will ignore instances—instances UR accommodates—of an action A's being such that some self-forming action causally contributes to A's occurrence in a way that makes a difference to whether or not A occurs.[18] Incidentally, while Kane takes the notion of causal contribution as primitive, I do not and offer an account of it in the next chapter.[19]

So, how should we understand Kane's conception of a self-forming action? Recall Juan's migrating to Nogales rather than staying in Huachichil. According to Kane, since Juan is personally responsible for migrating to Nogales, we should be able to trace the chain of causal contribution back into the past until we *stop* at an event for which Juan is personally responsible. For example, we may suppose that Juan reached Nogales because he rode on a grain truck; he rode on a grain truck because he hitch-hiked; he hitch-hiked because he walked to the highway

---

[16] Kane uses the terms 'willingly' and 'voluntarily' synonymously.

[17] Kane lifts the term 'arche' from the works of Aristotle. Kane (1996a: 33-4) states, "In addition to passages asserting alternative possibilities, according to [Aristotelian scholar Richard] Sorabji [1980: 233-4], there are others in which Aristotle asserts that 'the concept of an action being up to us is connected…with the concept of our being, or having within us, the "origin" (*arche*) of the action.' This second condition puts the emphasis for being up to us not on the power to do otherwise, but on the *source* or *explanation* of the action that is actually performed; that source must be 'in us.'"

[18] Kane (1996a: 221 note 11) says,
The voluntary doing or omission 'made a difference to whether or not E occurred' means that it played an essential role in the actual causal sequence leading to E—had this causal sequence not included it, the sequence would not have led to E or would have been less likely to have led to E.

[19] Kane (1996a: 28, 35, 221 note 11, 225 note 19) takes as primitive the relation of causal contribution. Hence, even if causal indeterminism is preferable to agent-causation, Kane could achieve additional rigor by grafting my general account of causal contribution proper onto his theory.

from his village of Huachichil, showing an outstretched thumb to passing vehicles; and he walked to the highway because…—and this backtracking must stop somewhere—…because he voluntarily makes the choice to migrate to Nogales instead of staying in Huachichil.

For Kane, voluntarily choosing to migrate to Nogales is supposed to count as Juan's self-forming action. Notice too that Juan's initial choice to migrate is intentional and therefore teleological. He has the goal of getting to Nogales. From the moment of his choice, he intends to get to Nogales. Kane (1996a: 78) states, "We can now say that *the meaning of 'ultimate' in [my definition of free action] is supplied by UR*: to ultimately create and sustain one's ends or purposes is to create and sustain intentions by acts of will (choice, decisions, judgments and efforts) for which one is ultimately responsible in the sense of UR."[20]

The notion of an act's being voluntarily performed is crucial to Kane's theory. According to Kane (1996a: 30), Juan's act of choosing to migrate is *voluntarily* performed precisely because he chooses for reasons and his choosing is the result of neither coercion nor compulsion. Note that this requirement confirms the lesson we learned from our discussion of Frankfurt in Chapter 2. If Black makes Jones choose to kill Smith, Jones's choice to kill Smith is coerced or compelled. Regarding acting *for* a reason, Kane (1996a: 28) *qua* causalist asserts,

> Of course, to be 'correctly' or truly cited as reasons, psychological attitudes must play a role in the etiology of choice or action—they must influence choice or action in some manner or other that is not easy to specify.[21]

Hence, since Juan's choice is neither coerced nor compelled but is rather produced by DB, i.e., his desire-belief complex that favors his choosing to migrate, Juan voluntarily chooses to migrate.

For sake of argument, let's assume that Juan is personally responsible for choosing to migrate to Nogales. We assume, then, that sub-condition R of UR is met. Thus, at the very moment Juan voluntarily chooses to migrate, Juan could have done otherwise at least in the minimal sense of having multiple causally unclosed futures. Since we are assuming incompatibilism, at the very least, Juan's voluntarily doing otherwise (i.e., his voluntarily choosing to remain in Huachichil) is not causally closed at the very moment Juan actually voluntarily chooses to migrate. Since Juan's voluntarily choosing to migrate is *directly* free, it must meet the ultimacy sub-condition U of UR.

The remainder of this section addresses two important questions that naturally arise:

- Does Juan's self-forming action of voluntarily choosing to migrate satisfy sub-condition U of UR?

---

[20] Notice that this quotation suggests that UR is also sufficient for origination.

[21] Reasons causing an action is not nearly sufficient for the agent's acting for those reasons. Kane (1996a: 136) notes,

> But even the staunchest causalists do not believe that causal influence is *sufficient* for saying the agents 'choose for reason.' There is also a teleological element involved in choosing for reasons. The choices must connect the reasons in a certain way to future actions chosen and intended.

- Is the satisfaction of UR enough, intuitively, to capture Juan's originating his choosing to migrate?

After I explain how Kane's theory has the resources to answer the first question affirmatively, I argue that it is far from clear how UR even begins to explain how Juan is an ultimate creator or underived originator of anything included by his voluntarily choosing to migrate.

5.3.2   Do any actions meet Kane's UR?

If Juan's choosing to migrate satisfies U, then UR is met. So, how is Juan's choosing to migrate supposed to fulfill U? We need to get a better grasp of U. Kane (1996a: 36-7) recognizes that

> the backtracking condition U is the pivotal (and most problematic) condition of UR… For one thing…the U (or Ultimacy) part of UR invites a…regress: for *any* X and Y, if you are personally responsible for X, and if Y is an *arche* or sufficient ground of X, then you must also be personally responsible for Y.

At least for finite creatures like us, the regress U invites must either be non-vicious or else stop somewhere. If the regress stops, it stops at Juan's voluntarily choosing to migrate. The reason the backtracking is non-vicious or stops at the voluntary choice lies in Kane's technical concept of an *arche* or sufficient ground.

So, what sorts of thing count as *archai* or sufficient grounds? Kane (1996a: 73) stipulates that they come in three varieties:

> They may be (i) (logically) sufficient *conditions*, (ii) sufficient *causes* (i.e., antecedent circumstances plus laws of nature) or (iii) sufficient *motives*. The first two entail the existence of that for which they are reasons. The third, sufficient motive, applies to actions and is sufficient in the sense that, given the motive, the action for which it is a motive would be performed voluntarily and would not be omitted voluntarily.

Let's take these in turn. Let 'E' denote the event *Juan's voluntarily choosing to migrate to Nogales*. By hypothesis, Juan is personally responsible for E. Now consider (i). E has infinitely many logically sufficient conditions. For example, E is a logically sufficient condition of E. Of course, it is trivial that if Juan is personally responsible for E, then he is personally responsible for *this* sufficient ground. If there is any regress here, it is certainly not vicious.

There are other logically sufficient conditions for E. Here are three candidates. First, consider disjunctive events of the form *its coming to pass either that E occurs or that P*, where 'P' denotes some false proposition such as *the moon is composed of cheese, 2+4=9, Juan is already in Nogales*, and so forth. Second, consider conjunctive events of the form *its coming to pass both that E occurs and that Q*, where 'Q' denotes some necessary truth such as *2+4=6, Juan is not taller than he is, sound arguments are valid*, and the like. Third, consider certain epistemological events of the form *Subject S knows that E occurs*.

So, the logically sufficient conditions of E are myriad. Moreover, it is not obviously wrong to think that Juan is personally responsible for any Z of the sorts outlined above in virtue of his being personally responsible for E. Nor is it obviously wrong to think that Juan is personally responsible for any Z* of the same sorts of logically sufficient conditions for any Z, and so on *ad infinitum*. Thus, even though there is a regress of logically sufficient conditions of E, the regress seems non-vicious, for Juan may be personally responsible for all of them.

Now consider (ii). Notice that (ii) implies a special case of (i) that is more controversial than the logically sufficient conditions just discussed. If E has a sufficient causal condition C, then there is a logically sufficient condition for E of the form *C occurs and the laws of nature are L*.[22] Sufficient causal conditions are what Kane has in mind by "sufficient *causes* (i.e., antecedent circumstances plus laws of nature)" in clause (ii). Sufficient causes, for Kane, are deterministic causes—recall his (1996a: 13) claim that "the first two [(i) and (ii)] entail the existence of that for which they are reasons." They cannot be indeterministic causes, since E has indeterministic causes (*viz.*, Juan's reasons that favor his migrating) and Kane must stop the regress at E. By hypothesis, E is Juan's first directly free act. Since E has no sufficient causes, then it is trivially the case that Juan is personally responsible for every sufficient cause of E.[23]

Finally, consider (iii). Since Juan's case meets UR, either E (i.e., Juan's act of voluntarily choosing to migrate) has no sufficient motive or Juan is personally responsible for every sufficient motive of E. Kane's theory entails, as I explain below, that there is no sufficient motive for E.

It is tempting to think that Juan has a sufficient motive for E. For, we've already assumed that Juan voluntarily chooses to migrate and that his reasons produce E. Since Juan's reasons that produce E include his wants/desires, and since Kane stipulates that motives just are wants/desires, Juan has a motive for E. According to Kane (1996a: 30), an agent's motive moves him to act only if the motive is a part of the etiology of the act. However, we also assume that Juan is not personally responsible for any antecedent causes of E, since E is supposed to be the action that stops the regress. That is, since E is supposed to be directly free, Juan cannot be personally responsible for any motives producing it.

This tempting line of thought underscores two important points. First, it follows that E satisfies UR only if there is no sufficient motive for E. So, Kane's account should imply that E has no sufficient motive. Recall Kane's (1996a: 73) claim that a motive "is sufficient in the sense that, given the motive, the action for which it is a motive would be performed voluntarily and would not be omitted voluntarily." This leads to the second point: it seems to follow that whether or not Juan has a sufficient motive for E boils down to whether or not Kane's (1996a: 73) clause "would not be omitted voluntarily" applies in Juan's case. For, Juan has a motive for choosing to migrate, and Juan's act of choosing to migrate is performed voluntarily.

---

[22] Recall from §3.1.1 that *C* is a sufficient causal condition of *E iff C* occurs, *C* does not imply *E*, and it is a law of nature that if *C* occurs at a time *T*, then *E* occurs at or later than *T*.

[23] Incidentally, clause (ii) bears most of the load of Kane's (1996a: 74) claim that UR entails incompatibilism. For, if determinism is true, then there are sufficient causal conditions (i.e., sufficient causes) for all of Juan's actions and they occur long before Juan is born. However, Juan, though he is personally responsible for some of his acts, cannot be personally responsible for these ancient sufficient causes, since there is nothing Juan could ever voluntarily do that would cause whether or not these sufficient causes occur—see Kane (1996a: 35-7). I assume with Kane that backward causation is impossible. Event E cannot cause some earlier event E*.

From the surface grammar of Kane's (1996a: 73) entire claim, one might be tempted to parse "would not be omitted voluntarily" with the proposition

(1)   Juan's choosing to migrate would not be omitted voluntarily.

If (1) is true, then Juan's choosing to migrate has a sufficient motive, which implies there is a sufficient ground—*viz.*, Juan's motive for choosing to migrate—for which Juan is not personally responsible. Hence, if (1) is true, then Juan's case does not satisfy UR.

At first blush, it appears that (1) is true. After all, since Juan actually does choose to migrate, he does not omit his choosing to migrate. *A fortiori* Juan does not omit his choosing to migrate voluntarily. However, the subjunctive nature of (1) suggests a more charitable reading, such as the following counterfactual:

(2)   Were Juan to omit his act of choosing to migrate, then Juan's choosing to migrate would not be omitted voluntarily.

To simplify the example by putting it into positive terminology consistent with how we've set up Juan's case, Juan would omit his choosing to migrate in virtue of making the alternative choice to remain in Huachichil. Hence, the truth of (2) rises or falls with the proposition

(3)   Were Juan to choose to remain in Huachichil instead, then Juan would not voluntarily choose to remain in Huachichil.[24]

Kane's later discussion of sufficient motives confirms our choice of (3) as the crucial condition required by an action's having a sufficient motive. He (1996a: 113, <u>my emphasis</u>) writes:

> Sufficient motives are sets of reasons or motives (wants, beliefs, etc.) that explain actions. An agent has a sufficient motive for doing something A at a time t <u>when</u> the agent's reasons or motives are such that, given them, if the agent did A at t, it would be *voluntarily* [i.e., causally influenced by those reasons[25]], and if the agent did other than A at t, it would <u>not</u> be voluntarily.

This is the most detailed explanation Kane proffers for his technical concept of a sufficient motive. But even here Kane only gives a sufficient condition for being a sufficient motive. In order to show that Juan's choosing to migrate does not have a sufficient motive, Kane should be providing a necessary condition that Juan's case intuitively fails to meet. Since Kane nowhere else offers such a necessary condition, I recommend that we understand Kane's use of 'when' in its definitional sense. Consequently, I extract the following precise conditions for a motive's being sufficient:

---

[24] So as not to be bogged down unnecessarily, I suggest that the alternative action B is Juan's choosing to migrate to Nogales.

[25] See Kane (1996a: 28).

(SM)   Agent Δ has a *sufficient* motive M for action A *iff* Δ has a motive M for performing A; were Δ to perform A, A would be performed voluntarily; and *for any different action B not causally closed by Δ's complete circumstances*, if Δ were to perform B instead, then Δ would not perform B voluntarily.[26]

Notice that having a sufficient motive for an action A does not entail that A occurs; otherwise, we'd have another special case of a (logically) sufficient condition.

We're now in a position to see why Kane would think E has no sufficient motive. By hypothesis, Juan's situation meets every other condition in UR. Juan voluntarily performs E and so has a motive for performing E—the question is whether this motive is a sufficient motive. Also, by R, Juan could have voluntarily done something else. Choosing to remain in Huachichil nicely fits this bill. And were Juan to perform this different action instead of E, this different action would be his voluntarily choosing to remain in Huachichil. It follows that it is not the case that were Juan to perform B, he would not voluntarily perform B. (3) is false. So, according to (SM), though Juan has a motive for E, Juan does not have a sufficient motive for E. Thus, it is trivially the case that Juan is personally responsible for every sufficient motive for E. In conclusion, since (i)-(iii) exhaust the sorts of potential sufficient ground, Juan is personally responsible for every sufficient ground of E.

Moreover, Juan's case is arbitrary. In general, then, (SM) guarantees that the responsibility portion R of Kane's (UR) entails that there are no sufficient motives for directly free actions. Hence, if R is satisfied, the only really interesting aspect of the ultimacy condition U for an agent's being personally responsible for a directly free action A is that the agent is personally responsible for every sufficient causal condition of A.[27] So, for the causal indeterminist, the ultimacy portion U of UR is trivially satisfied by an agent's directly free action simply in virtue of having no deterministic cause.

5.3.3   Is UR Enough for Ultimate Origination?
    Let's assume, then, that there are some actions for which an agent is ultimately responsible in the sense outlined by UR. Is UR enough to satisfy the origination condition that free action requires? Kane seems to think so, but this section explains that UR is not enough.

---

[26] One might be tempted to argue against (SM) as follows. Suppose that M is a sufficient causal condition of Δ's action A. Intuitively, M should count as a sufficient motive. However, Kane's definition permits M not to count as a sufficient motive. E.g., even though M deterministically causes A, M is not a sufficient motive if it happens to be the case that if Δ were to perform alternative action B instead, then Δ's performing B would be coerced or compelled. The objection is not telling against Kane's overall theory. For, even though M does not count as a sufficient motive in the objector's example, M still counts as a sufficient ground for A in virtue of being a sufficient cause in Kane's stipulative sense of 'sufficient cause'. Again, see Kane (1996a: 73). So, Kane has his bases covered. Moreover, Kane may easily amend (SM) by disjoining "M is a sufficient causal condition of A" to the explanans.

[27] I suspect that some compatibilists would see this result as interesting in that it would be their primary reason to reject his theory. Again, one of the reasons Kane (1996a: 59, 77) advances UR is to provide a foundation for denying compatibilism that is better than typical versions of the Consequence Argument.

Since UR provides only necessary conditions for ultimacy, Kane perhaps intends to explain origination by buttressing UR with other concepts (such as his notion of a self-network). I conclude my discussion of Kane by arguing that he offers no good reason to think his complete view implies the origination condition. My discussion of Kane, together with Whenceness, motivates a general argument against any causal indeterminist view.

In the wake of his treatment of UR, Kane (1996a: 79) outlines what he calls the *four conditions of sole authorship*. He introduces his conditions of sole authorship by stating, "The image of a kind of freedom that would satisfy UR is a remarkable one, but it is one that I think most of us have had at one time or another." This strongly suggests that UR implies these conditions of sole authorship. Moreover, immediately after discussing the conditions of sole authorship, he (1996a: 79) concludes, "These four conditions describe a kind of 'sole authorship' or 'underived origination' that many ordinary persons believe they want when they want free will." Therefore, it seems that UR is supposed to imply origination via implying the four conditions of sole authorship.

So there are two important questions. First, does UR imply the four conditions of sole authorship? Second, do the four conditions of sole authorship imply origination? After close reflection on his four conditions of sole authorship, one can see that UR fails to imply the conditions of sole authorship. Let's investigate his four conditions of sole authorship (in bold).

Kane's (1996a: 79) first condition reads, "(i) The source or ground (*arche*) of [a free] action would be in the agent or self, and not outside the agent." Now consider again our working example of a directly free action: Juan's voluntarily performing E. Recall that a logically sufficient condition for E is E itself. While I find it somewhat difficult to grasp precisely what it means for an event or state to occur *in* an agent, I don't find it unreasonable to think that E is, in some sense of 'in,' in (or internal to) Juan. However, E, though perhaps a sufficient ground in Kane's technical sense, can neither originate nor be a source of itself. And Juan does not originate E solely in virtue of E's occurring. Simple indeterminists might think so, but we've already seen that SI does not meet the origination condition. E is not intrinsically an event consisting of Juan's originating some event causally relevant to his migrating to Nogales. Thus, E does not meet Whenceness.

What other source of E might occur in Juan? Another logically sufficient condition of E takes the form *subject S knows that E occurs*. Surely the fact that E occurs is knowable—just let 'S' refer to God. According to Kane's account, the event *God's knowing that E occurs* is an *arche* of E. In what sense, though, is the event *God's knowing that E occurs in* Juan? It seems clear that this event is not in Juan. Thus, it's so far unclear how Kane's first condition applies to Juan's situation, and given the possibility of certain epistemic events that count as sufficient grounds for E, it seems clear that Kane's first condition of sole authorship is false.

Nonetheless, there are infinitely many logically sufficient conditions for E, and Kane uses the definite article, saying "(i) The source or ground…" So perhaps Kane has a more perspicuous and unique candidate for the source of Juan's performing E. Unfortunately, Kane offers no candidates.

Moreover, things don't get any clearer when we try to grasp what source or ground (*arche*) of E could be in Juan. For Kane's UR implies that there are no candidates. A sufficient causal condition would count as a source, but E has none. Sufficient motives would perhaps also count, but E has none of those either. It seems, then, that Kane is right in claiming that the source or ground (*arche*) of E would not be outside Juan—modulo apparently orthogonal events

such as *God's knowing that E occurs*. However, Kane offers no grounds for thinking that any source of E *is* in Juan.

Perhaps there's a clue to be found in Kane's implicit counterfactual "the source…would be in the agent." But what is the antecedent here? Is Kane's claim "*If there were a source (arche), then the source would be in the agent?*" If so, then since there is no *arche*, Kane's expression is vacuously true, albeit rhetorically powerful. So, on pain of interpreting Kane as speaking the truth, he's said nothing informative, for he has yet to provide a reason for thinking that there is an underived source of anything implied by E. *A fortiori* Kane has yet to provide a reason for thinking that Juan's situation implies that *he* is an underived source of any element relevant to his migrating to Nogales. E does not imply that Juan is an ultimate source of some event, nor is there a source for E that illuminates how Juan originates E.

Perhaps Kane's second condition of sole authorship helps, as it elaborates on the first. Kane (1996a: 79) states:

> [Condition (i)] would mean that (ii) if we were to trace the causal or explanatory chains of action backward to their sources, they would terminate in actions that can only and finally be explained in terms of the agent's voluntarily or willingly performing them (i.e., in self-forming actions or SFAs).

There is nothing here that explains how Juan is an underived source of anything implied by his directly free action, E. E just is Juan's self-forming action. E can neither explain nor be a source of itself.

Kane (1996a: 79) continues to the third condition of sole authorship, asserting, "(iii) The agent would be the sole author or underived originator of these self-forming actions and would thereby be ultimately responsible to some degree for the self which was formed by them and for subsequent actions issuing from that self." The first conjunct of this claim, of course, is the key. This third condition intuitively implies the condition of origination that ordinary people believe they want when they want free will. However, Kane does not explain how UR implies (iii). Nor is it obvious that UR implies (iii). At this point, (iii) is a mere declaration of faith. Kane provides no explanation for how Juan is the sole author or underived originator of his self-forming action E. Given Kane's CI, it is hard to see how Juan himself originates E, derivatively or directly.

Kane (1996a: 79) continues, "(iv) These self-forming actions would not be determined by anything within or outside the self for which the agent was in no way responsible." Strictly speaking, this is correct but unenlightening. For, self-forming actions are for the causal indeterminist determined by nothing, *a fortiori* they are not determined by anything for which the agent is not responsible. However, there is no account for how the agent actually originates something.

So, UR doesn't yield the origination free action requires. Again, to be fair to Kane, UR offers only necessary conditions for ultimate responsibility. Perhaps ultimate responsibility does imply origination. Perhaps not. The notion of ultimate responsibility is Kane's technical notion, and Kane nowhere else offers sufficient conditions for ultimate responsibility. At this point, then, we must consider the rest of Kane's theory. Perhaps Kane can meet the origination condition either without ultimate responsibility or with other conditions that when coupled with UR imply origination. What other resources does Kane have at his disposal?

5.3.4 Kane's Self-Network Falls Short of Origination

The only other place Kane could reasonably be interpreted as explaining how the agent originates her free activity lies in another technical notion, the notion of an agent's *self-network*.[28] Kane (1996a: 139) would explain Juan's self-network as follows:

> When we say that the conflicting reasons sets DB and DB* motivating Juan are both *his*—parts of his general motivational system—this corresponds, in neural terms, to saying that the neural connections representing DB and DB* are embedded in a more comprehensive network of neural connections representing the general motivational system in terms of which she defines herself as agent and practical reasoner.

Juan's self-network (SN) just is this comprehensive network representing his general motivational system. A self-network consists of an agent's psychological features—her plans, aspirations and hopes, ideals, desires, overarching intentions, and so forth. Kane (1996a: 140) claims that the intricate and complicated events constituting SN "push one competing reason-network DB over the top, so to speak, so that E is chosen for reasons DB rather than B for reasons DB* (or vice versa)—thus supporting the belief that the choice is Juan's doing, the product of himself."[29] Kane (1996a: 141) similarly asserts that

> the patterns of the self-network…push over the top, thus triggering the choice outcome. Whichever outcome is pushed over the top, however, would be the product of the self-network whose distinctive oscillations would be crucial to its coming about.

Is the Self-Network Enough for Origination? No, the operations that are an agent's SN do not obviously imply that the agent originates an event implied by her directly free activity.

Given that the self-network pushes over the top or triggers one choice, I think it is safe to say that the events consisting of Juan's SN cause E—abbreviated by *SN's causing E*. However, I fail to see how this supports the belief that E is the product of Juan himself. For, Juan is not identical to his self-network. Rather, Juan's having certain properties is a self-network. Speaking more loosely, Juan has a self-network.

If some SN were identical to Juan, then perhaps Kane could account for origination. Suppose Juan = SN. There is a problem here. First, free actions of finite creatures do not obviously entail a reductionist stance on what free creatures essentially are. To the degree to which one believes there are free actions and yet eschews the thesis that finite free creatures are essentially identical to some bundle of events (such as an SN) is the degree to which one should reject Kane's CI.

Whether or not Juan = SN, there is still the causally complex event *SN's causing E*. But *SN's causing E* is neither an event Juan obviously originates nor an event implying Juan's originating some event. It should be clear, given CI, that Juan does not originate *SN's causing E*. Thus, *SN's causing E* is not the event that satisfies Whenceness.

---

[28] Kane (1996a: 139-141). See also Kane (2002b: 424).

[29] I've altered the quotation just enough to apply to Juan's situation.

But why is *SN's causing E* not an event essentially consisting of Juan's originating some event? Suppose there is a very close self-network SN* in a deterministic system. Kane welcomes this possibility. He (1996a: 140) asserts that his characterization of the self-network's essence is "…meant to apply to mental action and agency in general, whether deterministic or nondeterministic, compatibilist or incompatibilist." So, the notion of a self-network does not preclude determinism. But notice that a determined self-network SN* yields no underived origination. If there is a sufficient causal condition X of *SN*'s causing E*, then it hardly seems appropriate to think that *SN*'s causing E* is a case of the agent's originating her activity. X might be a candidate for originating *SN's causing E*, but Juan is not. Thus, since there is no difference between *SN*'s causing E* and *SN's causing E* (except what are the laws of nature) and since there is no underived origination in the former case, there is no underived origination in the latter case.

Kane would agree that *SN*'s causing E* is not a case of Juan's originating some event. I take it, though, that he would dispute my contention that there is no difference between *SN*'s causing E* and *SN's causing E*. But how could Kane explain how there could be origination in *SN's causing E* but not in *SN*'s causing E*? The only difference lies in a lack of determinism in the former case. But this offers no positive explanation for how the agent originates something relevant to her activity. Randolph Clarke (2003: 106) observes:

> The self-determination, ultimacy, origination, or initiation present with the second [deterministic] agent differs from that present with the first [undetermined agent] only by an absence: there is absent, with the second agent, other-determination, but there is not present any further sort of positive self-determination, any further exercise of a positive power to determine what one does.

Since E is supposedly directly free, it seems that something in the neighborhood of E should imply that the agent originates some event implied by E. For instance, if the occurrence of E itself is not essentially an instance of the agent's originating some event, then it must be shown that the agent somehow originates E. But Kane has not shown this. With the exception that the agent must act voluntarily, Kane has only specified conditions that preclude origination by something else, something other than the agent. The agent must neither be coerced, nor compelled, nor determined by her complete circumstances. Since acting voluntarily does not strictly imply underived origination, it's hard to see how acting voluntarily plus these other negative constraints should imply underived origination.

5.3.5   The *Argument from Origination*

Clarke's recipe against Kane's CI has been rather simple. Imagine certain cases illustrating that an agent's being antecedently determined precludes him from originating his first directly free action. However, simply supplanting the deterministic causal relation obtaining between his having reasons and his action with an indeterministic causal relation obtaining between the same relata does not obviously imply that the agent is an underived originator of something relevant to his action. In each case the agent's putatively directly free act is directly causally produced by circumstances none of which implies that the agent is an underived source of his act. Susan Wolf's (1990: 10) comment, "The agents in these cases seem to be mere vehicles of change in the world rather than initiators of it," seems to apply. Ted Honderich

(1988: 387, *my emphasis*) expresses poignantly this common intuition underlying discussions on free will:

> Our thinking and speaking with respect to our selves, further and fundamentally, does not imply that they are merely items in standard causal sequences, caused causes of actions. My thinking and speaking about my self does not imply that it is merely a *link*. Talk of initiating, making, giving rise to, bringing about, producing, and so on implies, rather, an activity that is *not itself a product*. Above all, such an activity is implied by the idea that it secures to us open futures. More than that, our various expressions do positively suggest that an action, in being owed to an initiating self, is not owed to a standard causal sequence, or rather, to one which has the initiating self within it as an effect. The expressions do positively assert the self, *an active source* of the action.

When comparing the antecedently determined agent with a nearby undetermined agent, the only difference being a lack of determinism by the complete circumstances at the moment of the agent performs the action, it seems reasonable to think that the lack of circumstantial determinism in the latter does not *ipso facto* endow the agent with the feature *being an ultimate creator or originator*. An agent's originating his activity seems rather to consist in some positive feature she possesses, not some negative property such as being not determined by the complete circumstances in which the agent acts. Kane seems to have little resources in satisfying Whenceness. To review:

(Whenceness) Necessarily, for any agent Δ performing any directly free action A, Δ is an ultimate/underived originator of an essential element/part of A.

Here is my *Argument from Origination*. Clarke's recipe, Honderich's insight, and Whenceness suggest a general argument against CI. Let 'Δ' denote any finite agent. Let 'DB' denote the causes that directly produce *A*, where 'A' represents *Δ's voluntarily making some choice*. Consider the following events: *DB*, *DB's producing A*, and *A*. Note that according to CI, there are various other causes that in turn produce *DB* as well as *DB's producing A*. Now for the argument:

(6)   Causal indeterminism is correct only if *DB*, *DB's producing A*, or *A* are the only candidates with respect to Δ's essentially being an underived originator of an essential element of her freely acting—i.e., they are the only events that could satisfy Whenceness by strictly implying that Δ is essentially an underived originator of an essential element of her free action.[30]

(7)   Δ is not essentially an underived originator of DB, and DB does not strictly imply that Δ is essentially an underived originator of some event.

---

[30] The argument goes through just as well if we replace 'DB' with the event that is the agent's self-network 'SN.'

(8)   Δ is not essentially an underived originator of the causally complex event *DB's producing A*, and *DB's producing A* does not strictly imply that Δ is essentially an underived originator of some event.

(9)   Δ is not essentially an underived originator of A, and A does not strictly imply that Δ is essentially an underived originator of some event.

(10)   Thus, causal indeterminism fails to satisfy the origination condition free action requires.

    I take it that (8) and (9) are the most controversial premises. In particular, Kane would probably dispute (8)'s second conjunct and (9)'s first conjunct. I should think that Kane would deny (9) by claiming that Δ really is an underived originator of A. Why would Kane think this? Because Δ's very own belief-desire complex, DB, produces A. But is this a good reason? It seems that for this to be a good reason, *DB's producing A* should strictly imply that Δ is an underived originator of some event. In other words, it seems that what is really at issue is (8)'s second conjunct. Does *DB's producing A* strictly imply that Δ is an underived originator of some event? No.

    The reasoning behind (8) is straightforward. With regard to origination, there is no obvious and relevant difference between *DB's deterministically producing A* and *DB's indeterministically producing A*.[31] In the latter case, according to the causal indeterminist, DB produces A as much as DB produces A in the former case. The causal relation of production is the same. And production is production. Since *DB's indeterministically producing A* is just a <u>lack</u> of determination, it should not any more satisfy Whenceness than its deterministic counterpart. Moreover, its deterministic counterpart does not imply that Δ is an underived source of some event. So, the causal indeterminist needs to provide a positive account for how there <u>is</u> origination in the latter case but not in the former.

    As Randolph Clarke (2003: 101) notes,

> It is not strictly and literally true, on an event-causal account, that an agent is an originator of her actions, as the view allows that every cause of an action is itself caused by events that occurred long before the agent existed…

It seems that for an agent to originate some essential element of her free action, the agent must exercise some further causal power to make actual one of the alternatives left unclosed by previous events. To say that the relevant instance of indeterministic production yields origination whereas the corresponding case of deterministic production precludes origination and yet not say what motivates or accounts for the distinction seems inadequate a foundation for origination.

    Clarke is not idiosyncratic in thinking that the mere absence of antecedent determination by events in no principled way guarantees self-determination or origination. For example, Clarke (2003: 104-5) cites Paul Russell (1984: 169), who states,

---

[31] Of course, there may very well be a relevant difference here vis-à-vis the condition of dual ability. Nonetheless, the focus here is on origination, not dual power.

The [undetermined] agent…no more determines which set of incentives becomes his will in this case than in the [determined] case. That is to say, the opening up of possible actions in the [undetermined] case does not increase the agent's influence over what he does.

In addition, Gary Watson (1987: 165) asserts:

If we know of two agents, *A* and *B*, that both their behaviour is teleologically intelligible, and so on, then it is incredible to suppose that the additional information that determinism holds in *A*'s world but in *B*'s confers some special value or dignity to *B*'s life. For it implies no special *powers* for *B*. In whatever way *A* is supposed to be powerless, *B* is as well, if that is the only difference between them.

I conclude, then, that the Argument from Origination brings into relief CI's inadequacy in accounting for an agent's being an underived source of an essential part of her directly free action.

I anticipate that one might object that Whenceness is not precise enough to bear the load of an argument against both SI and CI. Defenders of SI and CI might demand that the agent-causalist provide the precise conditions for an agent's originating her activity. Then they could defend their views by challenging the definition or a premise in an argument for that definition (if there is any argument). I do not have a definition. This is a research program in its own right that I would like to take up in the near future.[32] I'm convinced, though, that the notion of ultimate origination has intuitive appeal. Consider Aristotle (NE II.5) who claims, "The voluntary is 'in' the agent, in contrast to acting under constraint 'to which the person contributes nothing.'" What does Aristotle mean by "contributes nothing," given that the agent acts? A puzzle arises. For, if an agent acts under constraint, then the agent acts. And if an agent acts, then the agent does something. And if an agent does something, then the agent contributes something. Thus, if an agent acts under constraint, then the agent contributes something. Therefore, it is not the case that a person acting under constraint contributes nothing. What should give? I think that a plausible way out of the puzzle is to recognize that Aristotle has a special sort of contribution in mind, one that requires that the agent contribute *qua* underived originator of an essential element of her action. I also refer the reader to Chapter 2, §2.2.2., where I discussed at length the intuitive nature of the condition of origination that free action requires.

Given that I have no rigorous definition of origination, my overarching strategy looks something like this. I have compared how various libertarians account for origination. I have

---

[32] Alvin Plantinga's (1993a, 1993b) use of the notion of proper function comes to mind. He does not give robust necessary and sufficient conditions; however, he exploits one's intuitive judgments about whether a cognitive faculty functions properly to undermine competitor theories of knowledge. It seems clear that he has shown that a belief is known only if the cognitive faculties functioned properly in the production of that belief. Another case that comes to mind is Peter van Inwagen's (1990b) using the concept of *life* to account for material composite objects. Van Inwagen, while discussing important features of the property *being alive*, does not give the precise conditions for its instantiation. Nonetheless, the notion of life does a great deal of work in his theory.

registered an intuition that some views fair worse than others. I shall in the sequel reflect on whether this intuition is defeated by considerations orthogonal to the issue of origination. All other things being equal, if there are two competing theories T and T* and, intuitively, T better captures how an agent is essentially an underived originator of an essential part of her directly free action, then one has a defeasible but good reason to accept T rather than T*. But perhaps all other things are not equal. Though respected philosophers hold to agent-causation (AC), detractors are vocal.[33] Just meeting Whenceness is insufficient for preferring AC over SI or CI as the best libertarian account of free action. For there may be insurmountable problems unique to AC, problems that are more damaging to AC than the failure to meet Whenceness is damaging to the SI and CI.

5.4   Agent-Causation 101

> *...it is true that human beings are the only putative examples we yet have of possessors of powers that are exercised at will. But human action is of great importance and interest to human beings, and if it can be understood only by employing a concept that applies to nothing else in their experience, they have no option but to employ it.*
> —*Alan Donagan (1989: 173)*

Commenting on AC, even Kane (1996a: 121) thinks that "…many philosophers believe that it alone captures basic incompatibilist intuitions about underived origination." I agree with these many philosophers. Alvin Plantinga (1990: 31) opines, "What is really at stake in this discussion is the notion of agent causation: the notion of a person as the ultimate source of action…." Kane (1996a: 238 note 28) also admits, "It is fair to say that most libertarians continue to be agent-causalists of one kind or another."[34] In the remainder of my book, I argue that most of these libertarians should, in their adherence to agent-causation, not be ashamed. This section provides a sketch of Timothy O'Connor's view of AC, noting some distinctive features. Throughout the final three chapters, I deepen and expand this agent-causal view as I illustrate how it survives various objections. The result will be a detailed and sophisticated agent-causal view that is both defensible and arguably the best libertarian view available.

---

[33] Following others, I use the hyphenated term 'agent-causation' instead of 'agent causation,' since nearly everyone believes that agents cause their actions. See Kane (1996a: 120), Broad (1952) and Chisholm (1995: 95). The question is whether every event of the contingent events implied by an agent's acting has among its causal contributors *only* other events.

[34] Proponents of some version of agent-causation, at one point or another, arguably include Aristotle [cited by Markosian 1999: 175], Plotinus, Francisco Suarez, Saint Thomas Aquinas, Duns Scotus, Samuel Clarke, Edmund Law, Thomas Reid, Immanuel Kant, C.A. Campbell, Richard Taylor, Roderick Chisholm, William Rowe, Alvin Plantinga, John Thorp, Timothy O'Connor, Randolph Clarke, Ned Markosian, Michael Zimmerman, J.P. Moreland, and God.

So, what are the basics of the agent-causal view?[35] *Agent*-causation, aptly enough, implies something about the nature of an agent. So, at the outset, the AC proponent is clear about what kind of thing an agent qua actor is. An agent is a substance. She is a continuant. She is an enduring entity having properties. The most salient of these properties are being rational and having understanding. An agent, when functioning properly, is capable of representing alternative courses of action to herself. She has desires and beliefs about those alternatives. An agent, to be clear, is not a property or bundle of properties. Rather, an agent has properties. A finite agent is not an event because an agent, while alive, is not herself something that occurs, and all events necessarily occur. Simple and causal indeterminists may wholly agree with this description of an agent.

The key to agent-*causation*, however, lies in which relation the agent bears to her directly free action. Van Inwagen (2000: 12) characterizes agent causation as

> a relation that agents—thinking or rational *substances*—bear to events. Agent causation is opposed to *event* causation, a relation that events bear to events. The friends of agent causation hold that the causes of some events are not (or are only partially) earlier events. They are rather substances—not *changes* in substances, which are of course events, but "the substances themselves."

Ginet (1997: 87) also describes the heart of agent-causation, saying that an agent acts freely

> only if there is a direct causal relation between the agent and the action (or some event internal to the action), a causal relation where the relatum on the cause side is not any event in or state of the agent but just the agent herself, that enduring entity.

The agent-causal relation is supposed to be a species of the efficient causal relation. Thus, AC implies that when an agent acts freely, there is some event having an efficient cause, where this efficient cause is strictly and literally the agent herself, an enduring substance.

Pacé Hume, the concept of causation is one of our most fundamental ideas. Thinking of more fundamental concepts with which one could analyze causation is nontrivial, difficult, and a task for enterprising philosophers. I will not proffer such an analysis. Nonetheless, in order to understand the nuances of the agent-causal position, especially as it compares to SI and CI, I take a small detour to note several relatively safe yet interesting assumptions about causation.

In the above discussion of SI and CI, we have taken 'causation' to mean 'efficient causation'. Intuitively, something's efficiently causing an event full-fledgedly makes the effected event occur. Efficient causes produce their effects.[36] Intuitively, whatever counts as *the* cause of an event is that which produces the event. I assume that one event's being a sufficient causal condition of another event is a clear example of something's producing an event, i.e., of something's being an efficient cause of an event.[37] Production, then, is a strong sense of 'causation'. But it is not the only sense.

---

[35] Hereafter, I grossly conflate nouns and adjectives, letting 'AC' ambiguously denote 'agent-causation,' 'agent-causal,' and the like. Context should make clear the intended use.

[36] See Anscombe (1998).

[37] For a definition of a sufficient causal condition, see Chapter 3, §3.1.1, or APPENDIX 1.

In addition to this strong sense of 'causation', there is a weak sense of 'causation'. The weak sense of 'causation' is simply that of causal contribution (or causal influence). Causal contribution is weaker than production in that some C may contribute causally to E even though C does not produce E. For example, one might ask, "What caused the leopard to attack the gazelle? Was it the gazelle's turning away to flee? Or was it the leopard's hunger?" Both, it seems. For, leopards only take down their prey from behind, never head-on. But satiated leopards don't attack either. So, neither the gazelle's turning away nor the leopard's hunger causes the leapord's attacking, in the strong sense of 'cause'. Rather, each seems to be only a *partial* or *contributing* cause. Indeed, the entire efficient cause invariably includes more factors in addition to these two contributing events. So, a clear case of C's contributing causally to an event E would be C's being an essential part of whatever produces E.

In sum, there are two senses of 'causing'—that of producing and that of causally contributing, where the former designates the strong sense and the latter designates the weak sense. It is important to note that whatever is caused in either sense of 'cause' is an event. That is, every effect is, we hereby stipulate, something that occurs. Let '$x \rightarrow y$' designate 'x produces the occurrence of y.'

Interestingly, when Goetz claims that, necessarily, every mental action is uncaused, he does not mean to say that an agent's mental act has no causal contributors. Mental acts occur in some sort of causally relevant context. They do not occur in a causal vacuum. For finite agents like us, a host of factors causally influence our mental actions, even if (as per Goetz's theory) no combination of these factors produces our mental actions. For example, my learning about the game of golf somehow and in some complicated way causally contributes to my choosing to play golf. My noticing that there are more eggs in the refrigerator causally contributes to my choosing to make an omelet. This is consistent with Goetz's claim that nothing produces my choosing to make an omelet. My climbing to the top of El Capitan causally contributes to my choosing to relax for a moment and enjoy the view. Intuitively, even if there is nothing that produces a choice, the choice is (or at least might be) made in a context of events, where each of these events causally contributes, at least in some small way, to the agent's choosing. If this is right, one might wonder whether Goetz would eschew an agent's reasons from at least contributing causally to her choice.

Equally interesting, when Kane claims that an agent's reasons efficiently cause her directly free action, it would seem that, strictly speaking, the agent's reasons are only part of what produces the action. Clearly there are events other than the agent's reasons that contribute causally to her directly free action, and it would be highly contentious to think that none of these is an essential part of whatever produces her action. I should think that Kane, were he more careful, would claim that an agent's reasons are essential parts of whatever produces the agent's directly free action.

I return now to my exposition of the agent-causal position. Pre-theoretically, nearly everyone believes that particular substances cause events. For example, it is easy to find people saying things like 'the brick caused the window to break,' 'the moon caused the tides,' and 'the mechanic caused the engine to turn over.' However, seeing how such claims may be merely elliptical is not difficult either. The propositions expressed may turn out to be 'the brick's striking the window caused the window's breaking,' 'the moon's gravitational attraction caused the tides,' and 'the mechanic's turning the key caused the engine's turning over.' If the latter class of expressions paraphrase the former class of expressions without loss of meaning, then there is good reason for thinking that a substance's producing some event is, strictly speaking,

the substance's having some feature's producing the event. And the latter is simply one event's producing another event.

There is nearly universal consensus that some events produce other events, as illustrated in the paraphrases above.[38] Other examples are easy to imagine. Dropping a sugar cube into a glass of warm water produces the sugar cube's dissolving. Or, more carefully, dropping a sugar cube into a glass of warm water is an essential part of what produces the sugar cube's dissolving. Certain cerebral events make occur, they produce or are parts of what produces, certain involuntary musculature contractions. Reading the word 'snake', together with other circumstances, efficiently causes your thinking of a snake. Moreover, one of the most striking examples of efficient causation seems to be one event's being produced by its sufficient causal condition.

The AC theorist may welcome the claim that *some* events are produced by other events—but only some. As we saw above, van Inwagen and Ginet note that traditional AC theorists contend that there are some events that are not produced by other events but are rather directly produced by a substance, where this substance is a person.[39] Thus, an agent-causalist will deny the Humean or reductionist contention that causation can be reduced either to some sort of necessary connection or to some sort of counterfactual dependence between events. There can be no necessary connection or counterfactual dependence between an enduring, thinking substance and some event. O'Connor (1996: 144) says,

> [C]ausation by events or states of affairs may conform to sharp, generalizable patterns—and it may even fall out of one's account of such causation that there will be such patterns—but causation, for the nonreductionist, is not constituted by such patterns.

---

[38] Still, there are thinkers that reject causation by events. In personal conversation, Tom Flint tells me that some people think that event-causation is elliptical for substance-causation, not the other way around. Flint mentions Michael Loux and Alfred Freddoso. He says that Michael Loux, for example, once said that the more he [Loux] thinks about it, the more amazing he finds it that anyone can believe that an event could cause anything, since you need something that has power to cause and only substances have powers, not events. Corroborating, in personal correspondence Professor Loux (2005) says, "Tom has my puzzlement right. You can mention it in your dissertation, but I don't have anything more to say."

Professor Freddoso likewise rejects event-causation and takes this to be a typical Medieval view. See his (2001) manuscript 'Suarez on Metaphysical Inquiry, Efficient Causality, and Divine Action,' http://www.nd.edu/%7Eafreddos/courses/450/suarez%20ch%2002.pdf. Nonetheless, in personal correspondence Freddoso denies that causation by a substance requires that the substance is an intelligent agent. He (2005) says,

> The one view I detest as utterly wrong-headed is that agent causation is something peculiar to God and intelligent agents. As far as I'm concerned, this is as bad as—and even worse than—the claim that all causation is event-causation.

In light of the controversy over whether events can be causes, Flint (2005) recommends softening my claim to say, "Most twentieth-century, Anglo-American, non-Medieval, analytic philosophers think that events can cause other events."

[39] Say that x *directly* produces y *iff* x produces y and there is no z such that x contributes causally to z and z contributes causally to y. See §5.4.1.

So, the AC theorist is a causal realist.[40] Consider, then, the following characterization of the agent-causal relation:

(ACR) Agent Δ agent-causes an event E =df. Δ is an agent, E is an event, and Δ directly produces E.

According to the AC theorist, an agent produces an event not by virtue of changing in some way, even if, as things might happen to turn out, the agent is changing in some way at the moment she *qua* substance produces an event. It follows that an agent produces some event not by virtue of *being changed* in some way, which is exactly what we should expect in light of the above criticisms advanced against SI and CI. For, if only events produce all of the event constituting my action, then it seems that I would be just as much on the receiving end throughout the causal process as I would be on the initiating end, in which case *I* would not be an underived originator of my behavior. If my being changed in some way produces my action, then since something else is the source of my being changed in this way (in virtue of producing my being changed in this way), then I cannot be an ultimate or underived source of my action. This consequence—that *an agent's being changed in some way* does not produce his directly free action—underwrites the contention that only AC can meet the origination condition that free action requires.

Now recall that the libertarian's goal is to make intelligible a directly free action that is not included by the set of complete circumstances in which the agent acts. How are we to understand the relationship between the proposition *an agent directly causes some event E* and *the agent performs a directly free action A*? Commenting on AC, Randolph Clarke (1996a: 20) notes that "…sometimes the event that is directly caused by the agent is regarded as the action…, while in other cases the agent's causing this event is considered the action."

One may also regard both the event caused by the agent as well as the agent's causing an event as actions. Consider the latter event, *the agent's causing event E*, or *its coming to pass that the agent causes E*.[41] Call events of this form *agent-causal events (ACEs)*. Now, upon reflection, it seems fairly intuitive that an ACE should count as an action. O'Connor (1995b: 181) says that "what is most intimately my activity is the causal *initiation* of my behavior…" Reiterating, he (2000a: 51) states that "…my causal production of certain events internal to myself would seem to be my activity par excellence." So, O'Connor holds that an agent's

---

[40] See Chapter 3, §3.1.1.

[41] Some philosophers deny that *the agent's directly producing E* is itself an event. For example, Thomas Reid (1895: Essays 1 and 4) thinks that an event must be either a substance's coming into existence or a change that a substance undergoes. An agent's producing an event is not obviously a change that the agent undergoes, so it is not an event. This commitment apparently preempts the need for the AC proponent to give the conditions under which something produces (or does not produce) an ACE. For, if every event has a cause (and Reid takes this to be an obvious truth) and if there is an ACE, then there is a cause of the ACE. What could be the cause of an ACE? If an event produces an ACE, then the AC theorist seems to be in the same sinking boat as the causal indeterminist. If the agent himself agent-causes the ACE (and so agent-causes his agent-causing the ACE, and so on), then perhaps there is a problematic regress. I assume that Reid is wrong, taking a more finely grained view of events. *An agent's producing E* happens and occurs. Whatever happens or occurs is an event. So, *an agent's producing E* is an event.

directly producing some event just is the agent's directly free action.[42] The relationship between the two propositions *an agent directly produces some event E* and *the agent performs a directly free action A* is, on our view, that of identity.

Some AC theorists hold that for an agent's directly free action A, the agent directly produces A.[43] By the end of this chapter, it will become clear why I think that this *product view* of AC is a bad idea. Rather, an AC theorist should take a *component view* of AC, where the directly free action just is the agent's ACE.

Thus far my view implies that an agent produces some event, but I've said nothing substantive about the nature of this effected event. So, what is its nature? According to O'Connor (2000a: 72), "Agent causes bring about immediately *executive* states of intention to act in various ways." What does this mean?

Forget, for the moment, about *what* causes a state of intention and *how* a state of intention is caused. There should be nothing outlandish about an agent's simply having an immediately executive state of intention to act in some way. For example, suppose that Al wants to sip Lagavulin. He sees the glass of scotch on the table before him and has the usual beliefs that accompany such a situation. However, simply believing there is scotch present and wanting to sip the scotch is insufficient for intending to sip the scotch. Suppose, though, that it then comes to pass that Al has a state of intention to sip scotch straightaway, right here and now. His environment cooperates, his motor faculties function properly, and this state of intention directly and seamlessly issues in his sipping scotch. Al executes what his intention prompted. He does what he intends to do.

Along the way, Al performs other actions as well. Al reaches for the glass. He clutches the glass. He lifts the glass. He determinedly smells its smoky canopy, and so forth. It is not unreasonable to say that Al's state of intention to sip scotch endures past the moment it began. His intention guides and directs his overall behavior of sipping the scotch. O'Connor (2000a: 72n) says that

> in the case of an observable bodily movement such as [sipping scotch], my action consists of the causal relation I bear to the coming-to-be of the state of determinate intention to [sip scotch], plus the sequence of events that flows from that state.[44]

Let's build an obvious choice into Al's situation. Let's imagine that Al encounters two adult-beverages: the glass of Lagavulin and a glass of fine Merlot. He desires the Merlot. He

---

[42] I agree with O'Connor that of the candidates for Δ's directly free action, (Δ→e) is a better candidate than *e*. For, necessarily, if (Δ→e), then *e* occurs. However, *e* might occur without (Δ→e). So, if both *e* and (Δ→e) are actions, then *e* would not be a directly free action. Thus, if (Δ→e) occurs, then it seems to be the best candidate for Δ's directly free action. Moreover, to be clear, my interest concerns only directly free actions.

[43] For example, see Clarke (2003, 1996a, 1993).

[44] O'Connor's claim raises complex questions that need not concern us here. Presumably, O'Connor wants bodily actions to consist of these causal sequences that issue from intentions. But many sequences of events issue from any given intention. Which sequence is relevant? Where does the sequence end? How does one determine the point at which such a sequence ends? Since my focus is on directly free actions, these important questions need not concern us.

also wants to sip scotch. He knows he will only take one. But which one? He briefly deliberates, considering his reasons favoring each choice. We suppose that each course of action is not causally closed and so remains undetermined. He thinks to himself, "Last time I had the scotch, and it was delicious. However, Tom is raving about the Merlot, and I've yet to try this vintage." Suppose that Al then forms a state of intention to sip scotch straightaway, right then and there. And suppose that the example proceeds as it did before.

Now, finally, focus on what produces Al's intention. What made this state of intention (INT) occur? It is possible that certain events produce Al's INT. Possibly, Al's psychological constellation, consisting of certain beliefs and desires, produces his INT. Nevertheless, there may be something else that might produce his INT. Suppose that Al himself, the man, the substance, that solid individual, produces INT. Then, according to the AC theorist, the event *Al's directly producing INT* (i.e., Al → INT) just is Al's directly free action. We may, without doing damage to ordinary language, let the ACE, *Al → INT*, designate Al's decision. By causing one intention rather than another, Al ipso facto decides which intention to have. His directly free act just is his forming this intention. He freely decides.

Since my concern lies only with the intelligibility of directly free action, we need not worry about how precisely Al's executive intention causally issues in the remainder of his action(s). It is worth highlighting, though, that of all the candidates for the event-constituent of an ACE, O'Connor's selection of the agent's intention is smart. Intentions seem to occur at the leading edge of our free actions, and intentions are packed with explanatory value.[45]

---

[45] Kane (1996a: 25, 26) explains,

Intentions (i) explain and motivationally sustain intentional actions; (ii) they function as plans guiding and monitoring behavior; (iii) they help to coordinate agents' behavior over time and the interactions of the agents with other agents; (iv) they prompt, and play motivating roles in, practical reasoning; and (v) they appropriately terminate practical reasonings…[and] (vi) intentions express an agent's purposes or goals (which is a role that desires or wants alone do not perform, since what is desired or wanted is not always selected as a goal)…by virtue of the fact that contents of intentions describe these purposes or goals.

Kane borrows much of this from Mele (1992: 140) and finds inspiration for these ideas in Michael Bratman (1987), Myles Brand (1984), John Searle (1983), Gilbert Harman (1986), and Carlos Moya (1991).

If agent-causation is possible, there may be other candidates for what kind of event the agent directly produces. Perhaps the effected event is a choice, as Goetz (1997, 1988) would conceive it—modulo his assertion that choices cannot be produced. Or, see Avak Howsepian (1999), who seems to think that the agent directly produces a stopping of prior events from resulting in their characteristic effects, where this stopping *thereby* permits other, competing events to issue in their characteristic effects. For example, an agent deliberates about whether to A for reasons R or to A* for reasons R*. Presumably, reasons R and R* vie for the spot of causing their corresponding actions, A and A*. At some point, the agent agent-causes the stopping of R's having any effect, thereby permitting R* to issue in A*. So, the agent performs A*. On this view, instead of the agent's directly producing A* or any essential element of A*, she performs A* in virtue of somehow agent-causally blocking R from causing A. While interesting, it seems that the agent's directly producing the blocking event seems to be a doing *per se* and should have some intentional component essential to it. Special thanks to Tom Flint for bringing Howsepian's view to my attention.

So, on the AC view, an ACE is a directly free action. O'Connor (2000b: 113) summarizes our sketch of the AC view:

> According to some of us, [besides event-causation] there is another species of the causal genus, involving the characteristic activity of purposive free agents. Such agents can represent possible courses of action to themselves and have desires and beliefs concerning those alternatives. Against that background motivational framework, they themselves directly bring about immediately *executive* states of intention to act in various ways. This direct causing by agents of states of intention goes like this: As with mechanistic causes, the distinctive capacities of agent causes ('active powers') are grounded in a property or set of properties. So any agent having the relevant internal properties will *have it directly within his power to* cause any of a range of states of intention delimited by internal and external circumstances. However, these properties function differently in the associated causal process. Instead of being associated with direct causal functions from circumstances to effects, they (in conjunction with appropriate circumstances) *make [causally[46]] possible* the agent's producing an effect. These choice-enabling properties ground a different type of causal power or capacity— one that in suitable circumstances is freely exercised by the agent himself.

So what are the similarities and differences between AC, on the one hand, and SI and CI on the other hand? With the causal indeterminist but against the simple indeterminist, the AC proponent holds that making sense of an agent's self-determining activity, her directly free action requires causal elucidation. However, with the simple indeterminist, the AC advocate denies the causal indeterminist's assertion that an agent's reasons must causally produce her directly free action. Indeed, the agent-causalist believes that there need not be any event that produces an agent's directly free action. With the simple indeterminist, the AC theorist also denies that an agent's directly free action is the event *the agent's having reasons such-and-such's causing some event*.

It should now be rather obvious how the AC theorist addresses Whenceness, which states:

(Whenceness) Necessarily, for any agent Δ performing any directly free action A, Δ is an ultimate/underived originator of an essential element/part of A.

In Al's case, for example, the directly free action just is *Al's directly producing an intention to sip scotch straightaway*, i.e., *Al→INT*. The crucial issue, then, comes down to whether Al is essentially an underived originator of an essential element of his ACE.

Now either something directly produces *Al→INT*, or not. If we take the second horn, then clearly Al is an underived source of an element essential of his directly free action. For Al directly produces INT, and INT is an essential element of the directly free action, *Al→INT*. However, suppose we take the first horn. Suppose that some event directly produces *Al→INT*.

---

[46] I insert 'causally' since the modality here is certainly neither epistemic possibility nor that of broadly logical possibility. Intuitively, nothing broadly logically possible is made to be the case by anything else.

Then agent-causation seems not to meet Whenceness after all, as it would fall prey to the same sort of objection raised above against SI and CI. For, if an ACE could be directly produced by previous events (whether deterministically or indeterministically), then the agent would not be an underived source of an essential part of her action. Moreover, it would seem that there is no principled distinction between cases where an antecedent event produces an ACE and cases, if there are any such cases, where no event produces the ACE—including cases where the agent, and not some event, directly produces her own ACE.[47] In other words, if something could produce Al's ACE but it just so happens that nothing does, then even if Al happens to be an underived originator of an essential element of the ACE, Al's performing the ACE would not *strictly imply* that he is such an underived originator. It would be nice to have a reason for thinking that Al's directly free action, which in this case is his ACE, strictly implies that Al is an underived originator of an essential part of his ACE.

It is at this point where I think O'Connor advances the discussion considerably. He provides an argument or two for thinking that an ACE *could not* be directly produced by something.[48] It is vitally important to note that this is not to say that nothing could causally contribute to an ACE. Causal contribution, recall, is causation in the weak sense of 'cause'. The claim, rather, is that nothing could cause an ACE in the strong sense of 'cause'. For example, O'Connor is right only if there could not be any sufficient causal condition for an ACE.

In the next several sections, I also argue that nothing could produce an ACE. I provide at least four arguments for the same conclusion. While I found inspiration in O'Connor's work and general strategy, my arguments are a bit more detailed and rigorous. I conclude that the contention that nothing can produce an ACE has merit and considerably fills out the agent-causal view.

5.4.1   On Directly Producing an Agent-Causal Event (ACE)[49]

This section argues that nothing could directly produce an agent-causal event (ACE). I make use of the three operators '→', 'cc→', and '•→'. They are defined as follows:

- (x→y) *iff* x produces y

- (x cc→y) *iff* x causally contributes to y

---

[47] William Rowe (2000: 430 note 8) notices that Sir William Hamilton, in commenting on Reid's theory of agent-causation, expresses this point aptly,

> Only if he were not determined to that determination. But is the person an *original undetermined cause* of the determination of his will. If he be not, then is he not a *free* agent, and the scheme of Necessity is admitted.

[48] O'Connor (2002a: 136; 2000a: 53, 61; 1996: 147-8; 1995b: 186). To see O'Connor's arguments fully quoted, consult APPENDIX 2. William Rowe (2000: 430, 439, 445; 1995: 159) argues that Thomas Reid holds the same thesis, but the reasons he adduces on its behalf are altogether different than those discussed below. The fact that two of the most prominent contemporary AC theorists come to hold the same pivotal proposition lends credence to its being essential to an adequate agent-causal view.

[49] Special thanks to E.J. Coffman for helpful remarks on this section.

- (x•→y) *iff* x directly produces y

I've already discussed the intuitive distinction between production and causal contribution. The idea behind direct production is straightforward. Intuitively, x *directly* produces y just when x produces y but x does not in any way cause something else that in any way causes y. That is, *x's producing y* must in no way be causally mediated through some other event; rather, x immediately produces y. Say, then, that x *directly* produces y *iff* x produces y and there is no z such that x contributes causally to z and z contributes causally to y. That is, (x•→y) *iff* (x→y) and ~∃z{(x cc→z) & (z cc→y)}. I take it as an obvious truth that, necessarily, something produces an event only if there is something (perhaps something else) that *directly* produces that event. That is,

(11)   □∀x∀y{(x→y) ⊃ ∃z(z•→y)}.

Now, this section investigates whether or not producing an ACE is possible. With (11) in mind, the question concerns whether or not it is possible that something *directly* produces an agent's ACE. I shall argue that it is not possible. That is, I shall argue for the falsity of the following proposition:

(12)   ◊∃x∃y∃z{z•→(x•→y), where x is an agent}.

The pivotal premise of my overall argument is the claim that whatever directly produces a causally complex event essentially contributes causally to the cause-constituent (intuitively, the leading edge) of that causally complex event. More formally, my pivotal premise is

(13)   □∀x∀y∀z{(z•→(x•→y)) ⊃ (z cc→x)}.[50]

Calling (13) pivotal is not to say that (13) is controversial. On the contrary, (13) strikes me as self-evident as any substantive philosophical proposition could be. I know of no clear counterexamples to (13). Examples confirming (13) are fairly easy to imagine. O'Connor (2000a: 53) says, "Consider a familiar sequence of events. My finger presses a doorbell button, the doorbell rings, and your cat jumps in fright." Let 'Jumping' designate the event *your cat's jumping*. Now, suppose that some event directly produces Jumping, and let 'Ringing' designate this event that directly produces Jumping. Hence, we have the causally complex event *Ringing•→Jumping*. Let's make the intuitive, simplifying assumption that *the doorbell's ringing* is a part of Ringing.[51] Now suppose that there is something that produces (Ringing•→Jumping), designated by 'Pressing'. We may suppose that *your finger's pressing the doorbell button* is a part of Pressing. So, Pressing•→(Ringing•→Jumping). Now ask yourself, "Does Pressing contribute causally to Ringing?" It seems so. After all, Pressing contributes causally to *the doorbell's ringing*, which is a part of Ringing. Intuitively, Pressing contributes causally to

---

[50] Read as '□∀x∀y∀z(z directly produces *x's directly producing y* ⊃ z causally contributes to x).'

[51] Anytime I speak of a part of an event E, I assume that the part is itself some event E* or at least the cause-constituent of E. In other words, let '*x* is a part of event *y*' stand for 'either *y* is a causally complex event whose cause-constituent is *x* or else *x* is an event-constituent of *y*'.

Ringing in virtue of *your pressing the doorbell button*'s contributing causally to *the doorbell's ringing*. Hence, (13) is confirmed.

So (13) is a pivotal premise, yielding considerable conceptual mileage for the AC theorist. Again, I think (13) is self-evident or very close to it. Nevertheless, while *I think* (13) is obvious enough to be an underived premise in any good argument, I shall offer four individual arguments for believing (13). If I'm right and if (13) does not impinge upon anything the AC theorist already has good reason for believing (and I maintain that it doesn't), then the AC theorist may without embarrassment claim that (13) expresses a general truth about causation, in which case she may confidently use (13) as a premise.

5.4.2   The Argument from the First Instance

The first argument for (13) is bold, running as follows. We may speak <u>as if</u> some z could directly produce *x's directly producing y*, but this is just shorthand for saying that z directly produces x, where x in turn directly produces y—like unto dominoes falling. To account for the direct production of every causally simple event suffices for accounting for every instance of direct production simpliciter. Each producing a producing can be paraphrased without remainder in terms of something's producing a causally simple event. So, one should think that z directly produces (x•→y) *only in virtue of* directly producing x. But this is just to say that the relation between z and (x•→y) is not really that of direct production at all; rather, it is some other relation holding in virtue of the metaphysically basic relation of direct production.

The contention here is that (13) is vacuously true because its antecedent is necessarily false. There cannot be such a thing as directly producing a direct production, regardless of whether the direct production is an ACE or a causally complex event consisting of one event's causing another event. For, that which is directly produced is essentially a causally simple event—e.g., an object's having certain intrinsic properties. O'Connor (1996: 147) presses this point, saying:

> For instantiations of causal *relations* (causally complex events) are not themselves directly on the receiving end of other causal relations—instead, instantiations of intrinsic properties (causally simple *states* or events) are. Causings are the *producing* of events, rather than what are *produced* (in the first instance).

In sum, either this bold argument succeeds or fails. If the argument succeeds, then (13) is true. If it fails, which for the sake of argument I assume in the next three arguments, then there could be something that directly produces an instance of direct production.

5.4.3   The Argument from the Production of Essential Parts

Intuitively, producing an event requires causally influencing the effected event's parts—e.g., its event-constituents. (Again, anytime I speak of a part of an event E, I assume that the part is itself some event E* or at least the cause-constituent of E. I stipulatively understand the relation designated by '*x* is a part of event *y*' as 'either *y* is a causally complex event whose cause-constituent is *x* or else *x* is an event-constituent of *y*'.) Here is a limiting case. Suppose that X produces Y. Clearly, Y is an essential part of Y, though of course not a proper part. So, X at least causally influences Y. Here's another example. Suppose that X produces the event

Fred's speaking the sentence 'Naturalism is all the rage these days.' Clearly, X produces Fred's speaking this sentence only if X at least contributes causally to Fred's saying the word 'naturalism'. For, saying the word 'naturalism' is an essential part of saying the sentence 'Naturalism is all the rage these days.' Here's another example. Suppose that some X is a sufficient causal condition of *the football's spiraling toward the receiver*. Since every sufficient causal condition is a deterministic cause and every deterministic cause produces its effect, X produces *the football's spiraling toward the receiver*. It seems that we must admit that X at least causally influences *the football's spiraling*, since the latter event is intuitively an essential part of *the football's spiraling toward the receiver*. It seems, then, that the following principle (the Production of Essential Parts principle) is true:

(PEP) $\Box \forall x \forall y \forall z \{((x \bullet \rightarrow y) \ \& \ z$ is an essential part of $y) \supset (x \ cc \rightarrow z)\}$.

Principle PEP strictly implies (13). For conditional proof, assume $c \bullet \rightarrow (a \bullet \rightarrow b)$. An essential part of $(a \bullet \rightarrow b)$ is a. By PEP, $(c \ cc \rightarrow a)$. Thus, if PEP is true, then $c \bullet \rightarrow (a \bullet \rightarrow b)$ only if $(c \ cc \rightarrow a)$, which completes the conditional proof. Generalizing, PEP is true only if (13) is true. PEP is true. I conclude that (13) is true.[52]

5.4.4   The Argument from Deficiency

Suppose that x occurs and consider some z that in no way contributes causally to x. That is, suppose that x occurs and $\sim(z \ cc \rightarrow x)$. Suppose also that $(x \bullet \rightarrow y)$. Notice that $(x \bullet \rightarrow y)$ cannot occur unless x occurs. Clearly, something, something other than z, contributed to x in order to get the occurrence of $(x \bullet \rightarrow y)$. Since x must occur in order for $(x \bullet \rightarrow y)$ to occur, then if anything causes $(x \bullet \rightarrow y)$ in the strong sense of 'cause', z is not enough. Since z is in no way part of the etiology of x and x is required for $(x \bullet \rightarrow y)$ even to begin to occur, z is *deficient* when it comes to producing $(x \bullet \rightarrow y)$.[53] But this is just to say that z falls short of whatever it takes to produce $(x \bullet \rightarrow y)$. That is, since z does not causally influence x, z is not enough to produce $(x \bullet \rightarrow y)$. Again, something else is needed to contribute to x in order to get the occurrence of $(x \bullet \rightarrow y)$.

Therefore, given the assumption that x is an event, it follows that $\sim(z \ cc \rightarrow x)$ only if $\sim(z \bullet \rightarrow (x \bullet \rightarrow y))$. But notice that $\sim(z \ cc \rightarrow x)$ only if $\sim(z \bullet \rightarrow (x \bullet \rightarrow y))$ is just the contrapositive of (13). It follows that (13) is true at least in cases where the first quantifier ranges over events. Thus, barring any good reason for thinking that an agent-causalist is obliged to deny (13), the Argument from Deficiency provides additional support for thinking that (13) expresses a general truth about causation.

---

[52] One might be tempted to think that there could be cases where an event x causes *x's causing y*. Chisholm (1978: 625-6) once thought this. While I think PEP is more intuitive than and inconsistent with Chisholm's contention, see my Chapter 7, §7.4.2, for a more thorough criticism.

[53] That z is deficient when it comes to producing $(x \bullet \rightarrow y)$ is consistent with z's contributing causally to $(x \bullet \rightarrow y)$.

5.4.5 The Trigger Argument[54]

Suppose that (x•→y). Since (x•→y), x makes y occur precisely when y occurs. Directly producing y makes y occur right then and there. Something cannot produce y after y has already begun. Moreover, something cannot directly produce y without y's beginning to occur. So, when (x•→y), x gets y up and running, x makes y commence, x triggers y.[55] So, x directly produces y *now* only if x triggers y *now*, making y occur *now*.[56] More precisely,

(14)   □∀x∀y{(x•→y) ⊃ (x triggers y)}.

Intuitively, even if y takes a long time to finish occurring, since x triggers y, y at least has a first part to which x causally contributes. To get y up and running, as x does when it triggers y, y's first part must get up and running. And if x triggers y, making it occur precisely when it does, then x at least contributes causally to that first element of y. x triggers y only if x contributes causally to y's leading edge. I understate this point to minimize controversy. For, more controversially, it seems that x triggers y only if x triggers y's leading edge, i.e., only if x stimulates y's first part. But, less controversially, all we need to get (13) is the following:

(15)   □∀x∀y{(x triggers y) ⊃ ∃z((z is y's first part) & (x cc→z))}.

Here is a limiting case of (15). Suppose that x triggers y and that y is a temporally short lived event, occurring for only an instant and having no essential proper parts. Thus, the first part of y is y itself. Moreover, since x triggers y, x at least contributes causally to y. How else does something trigger an event if not initiating, stimulating, or influencing—at least in some small causal way—the leading edge of y? Thus, we confirm (15).

Obviously, (14) and (15) together imply

(16)   □∀x∀y{(x•→y) ⊃ ∃z((z is y's first part) & (x cc→z))}.

That (16) implies (13) is rather self-evident too. For conditional proof, suppose that c•→(a•→b). Given (16), it follows that ∃z((z is (a•→b)'s first part) & (c cc→ z)). What is the witness for this existential quantifier? The only candidate is *a*, the leading edge of (a•→b). Hence, (c cc→ a), which completes the conditional proof. Generalizing, (16) is true only if (13) is true. (16) is true. Hence, (13) is true.

In fine, I stand by (13) and put it to any critic of my agent-causal view to construct a clear counterexample against it.

---

[54] This argument bears a family resemblance to some of O'Connor's arguments cited *in extenso* in APPENDIX 2.

[55] Perhaps x triggers y in virtue of some z's triggering y, where z is a part of x.

[56] This is not to say that every part of y's etiology is part of what triggers y. For more on the notion of a trigger cause, see Fred Dretske (1988: 42).

129

5.4.6   Producing the Leading Edge Argument (PLEA)
    Given these four arguments, I conclude that we have excellent reason to think that (13) expresses a general truth about causation.[57] To review:

(13)   □∀x∀y∀z{(z•→(x•→y)) ⊃ (z cc→x)}.

    With a bit of reflection, one can see why nothing can produce an ACE, if (13) is true. For, it is simply impossible for something to causally contribute to *a substance*, as it is impossible for something to make a substance occur. A substance is not the sort of thing that can be made to occur.[58] And since there is no contributing causally to the occurrence of a substance at the moment of the substance's directly free action, there is no producing the cause-constituent (i.e., the leading edge) of the causally complex event *agent Δ's directly producing event e*. Thus, since (13) is true, nothing can produce an ACE. Hence, (12) is false.
    If one should want a more detailed argument, consider *the Producing the Leading Edge Argument (PLEA)*:

(17)   In an arbitrary world W, some *c* directly produces an arbitrary ACE, (Δ•→e), for some
        agent Δ. [For *reductio*, instance of (12)]

(13)   □∀x∀y∀z{(z•→(x•→y)) ⊃ (z cc→x)}. [Basic Conceptual Truth]

(18)   Thus, in W, *c* cc→Δ. [From 17,13]

(19)   ~◊∃z∃x(z cc→x, where x is a substance). [Basic Conceptual Truth]

(20)   Thus, (*c* cc→Δ) in W, yet ~◊(*c* cc→Δ), which is absurd. [From 18,19]

(21)   Thus, ~(17), i.e., nothing could produce an ACE. [Completes *reductio*, 17-20]

    Notice that these considerations mesh nicely with the thesis of incompatibilism. For, if freely acting requires agent-causation and an ACE cannot be produced by anything, then we can

_____

    [57] In personal conversation, Tom Flint raises a worry paraphrased as follows. Suppose that Screwtape has power over the causal laws connecting F and G. Suppose, though, Screwtape has no power over whether or not F occurs, yet he knows that F will occur. Nevertheless, he's able to make it the case that F produces G by modifying the laws. And suppose Screwtape does modify the laws so that F produces G. Then, intuitively, he caused *F's producing G* without contributing to F.
    There is a reply. Screwtape does not cause *F's producing G* in the strong sense of 'cause'. For all I've argued, something may directly cause—in the weak sense of 'cause'—a causally complex event without causally contributing to the cause-constituent of the complex event. Moreover, it is not at all obvious that Screwtape directly produces *F's producing G*, as he fiddles with the laws first. Thus, he doesn't directly produce *F's producing G*. For example, Screwtape is not able to create a sufficient causal condition of *F's producing G* without contributing causally to F; however, Screwtape may very well create a necessary causal condition of *F's producing G* without contributing causally to F.
    [58] Recall we are using '*x's producing y*' in the stipulative causal sense, where whatever is produced necessarily occurs.

see even more clearly why acting freely requires indeterminism. For, deterministic causes produce their effects, and since an ACE cannot be produced, there can be no deterministic cause of an ACE. Now this is not supposed to be an argument for incompatibilism. Perhaps there is an argument along these lines, but I will not pursue it here. I mention this result to provide further support for (13). The support, here, is an inference to the best explanation. (13), together with the theory of agent-causation, elucidates the folk's default conviction that a free action is essentially undetermined.

5.4.7 Addressing a Worry

So, I've proposed a detailed and multi-pronged argument against (12), i.e., I've proposed an argument concluding that nothing can produce an ACE. This argument spans §5.4.1 to §5.4.6. I've highlighted (13) as the pivotal premise in my argument against (12). Unfortunately, while I find (13) self-evident, others may not. My critic may object, "Does (13) really express a perfectly general truth about causation?" Self-evidence aside, my critic may press her skepticism further, claiming that my four arguments for (13) are compelling only if its quantifiers range over events. For example, the success of the Argument from Deficiency (§5.4.4) does not obviously strictly imply (13). Rather, it implies only a qualified version of (13), a version having (13)'s first quantifier range only over events. So the worry is that if (13) is used in an argument for thinking that nothing can directly produce an ACE, then (13)'s first quantifier must range over agents as well. For, the whole point of my overall argument is to reason that nothing can produce an ACE. If (13) expresses a perfectly general truth about causation, then (13) is true whether or not its quantifiers range only over events. My critic concludes that my overall argument is not very compelling.

I see two ways to respond to this worry. First, though my critic is right to point out that my Argument from Deficiency only attempts to earn the qualified version of (13), at least two of my remaining arguments for (13), if successful, earn the unqualified version of (13). For example, both the Argument from the First Instance (§5.4.2) as well as the Argument from the Production of Essential Parts (§5.4.3) are intended to go through whether their quantifiers' domains include only events or whether the domains include agents as well as events.

Second, the agent-causalist may take a more modest posture regarding her argument against (12), a posture that is still philosophically interesting. The AC theorist may hold that (13) is innocent until proven guilty. Why innocent? First, (13) has great intuitive appeal. Second, given the success of any of my four arguments, we know that there are no counterexamples against (13) that invoke cases of event-causation. Third, the truth of (13) together with the AC theorist's claim that a free action requires an ACE yields an inference to the best explanation for the folk's conviction that incompatibilism is correct.

The issue, then, is whether or not one may undermine my argument against (12) by providing a clear counterexample against (13). If my four arguments for (13) are compelling at least from an event-causal point of view, then the Eventist—i.e., someone believing that, necessarily, only events cause events—can provide no compelling reason for rejecting my arguments for (13). Thus, the issue hinges on whether anyone committed to the possibility of an ACE can provide a good reason for rejecting my argument against (12) by rejecting (13). Of course, my critic who believes in the possibility of an ACE can simply stipulate that something might directly produce an ACE, but this is just to assert the denial of my conclusion—i.e., it is

just to stipulate that (12) is true. It's hard, though, to see how this is a *principled* evaluation of my argument against (12).

I see only two plausible ways to undermine my argument against (12). First, one could proffer good reasons of some sort that pinpoints where all of my arguments for (13) go awry—reasons that do not stipulate that, possibly, something produces an ACE. Second, my critic could give a good argument for why an AC theorist is obliged to accept (12), which would ipso facto be a good reason to reject (13). Short of either one of these well-founded criticisms, I tentatively and confidently maintain that (12) is false. Nothing could produce an agent-causal event.

5.5   Agent-Causation Meets Whenceness: Comparisons and Contrasts

At this point, it should come as no surprise that Whenceness presents no difficulty for AC. Rather, AC flourishes. A moment of careful reflection shows that an ACE, which just is an agent's directly free action, strictly implies that the agent is an ultimate/underived originator of some essential element of her directly free action. An intelligent agent strictly and literally directly produces her intention to act in a certain way. A substance's strictly and literally producing an event is obviously a case of that substance's originating that event. And the origination here is essentially ultimate or underived, as nothing can produce it.

AC does not fall prey to the same sort of objection raised against Ginet's SI and Kane's CI. Since an ACE implies that its agent is an originator and since an ACE cannot be produced either by events or by an agent, there is no nearby world where the ACE occurs and its agent is obviously not such an originator. Hence, any given ACE strictly implies that its agent is an ultimate/underived originator of an essential element of her directly free action. Seeing how AC meets Whenceness perhaps captures the underlying hunch so many philosophers have had in finding AC the most attractive libertarian view.

Recall from §5.3.3 that Kane attempted to give four conditions of sole authorship, which if successful would apparently suffice for meeting Whenceness. I argued, however, that his conditions failed to capture sole authorship. The AC proponent, though, may supplant Kane's (1996a: 79) four conditions of sole authorship with four conditions of sole authorship a lá agent-causation:

(i)   An underived source of a derivatively free action is the agent herself, and a directly free action is intrinsically such that the agent is an underived source of her coming to have a state of intention to act in a certain way.

(ii)   If we were to trace the causal or explanatory chains of action backward to an underived source, they would terminate at the agent, where the first directly free action is the agent's producing a state of intention to act in a certain way, which is essential to the agent's free behavior (this directly free action is intuitively a self-forming action or SFA).

(iii)   These self-forming actions strictly imply that the agent is the sole author or underived originator of her intention to act and would thereby be ultimately responsible to some degree for forming the state of intention, for the state of intention itself, for the qualities of the self which is formed by the intention, and for subsequent actions issuing from her having these further qualities.

132

(iv)   Self-forming actions (i.e., ACEs) imply that the agent is a not-wholly-moved-mover.[59] An ACE is an agent's directly producing her intention. An ACE cannot be produced by anything. But an ACE may be causally influenced by events for which the agent is in no way responsible, which is just what we should expect for a finite agent who performs her first directly free action for which she is morally responsible in an environmental context.

Admittedly, of all of the libertarian views, our AC view is most similar to Goetz's SI. Goetz and I both claim that nothing can cause—in the strong sense of 'cause'—an agent's directly free action. Our reasons, however, are altogether different. Goetz claims that there can be no efficient cause of any mental action. Recall Goetz's (1997: 197) bold assertion, "An event which is efficiently caused is produced by that cause and as such is an occurrence with respect to which its subject is essentially passive." So, in addition to the claim that nothing can produce one's mental action that counts as free, Goetz holds the much more controversial claim producing *any* mental action is strictly impossible. Goetz takes this thesis as self-evident. For, as we saw, Goetz provides little by way of support for this broad and sweeping proposition. Finally, if Goetz cannot earn the claim that nothing could produce one's directly free action, then either Goetz must simply stipulate that this claim is part of his view or else fall prey to my argument against Ginet's SI.

In contrast to Goetz's SI, the AC proponent need not claim that nothing can produce a mental action, although this claim is consistent with the AC view. Moreover, for someone who already believes that one's directly free action is an ACE, he need not merely stipulate that nothing could produce a directly free action. For, the AC theorist has strong reasons for this controversial claim. The AC proponent, then, gives a positive and principled way to meet Whenceness. In contrast, it is far from clear how the simple indeterminist can do this—short of simply asserting without argument that every agent's mental action is essentially undetermined and is *ipso facto* an ultimate/underived originator of her mental action.

I conclude my discussion of the similarities and differences between AC and Goetz's SI by entertaining a likely objection to both of our theories. On both views, nothing and no one causes an agent's directly free action—in the strong sense of 'cause'. A fortiori, the agent herself does not produce her own directly free action. So consider my directly free action and ask, "Who or what produces it?" Goetz and I answer, "Nothing and no one." Some of our critics would likely claim that on the surface, this claim seems a bit jarring.[60] Shouldn't this give us pause? After all, shouldn't I be able to produce my own directly free action?

The AC theorist may concede the counterintuitive nature of the claim that nothing and no one produces anyone's free action. Nevertheless, there are considerations that suggest why it strikes us as counterintuitive. We pre-theoretically think that when one acts freely, one causes (produces) something. Yet this intuition is right on the mark! This pre-theoretic intuition confirms rather than disconfirms AC. The AC proponent admits that when acting freely, one produces something. Moreover, one's producing something just is one's directly free action. In sum, it's not as if there is no producing involved in any directly free action. *That* is a jarring

---

[59] The locution "not wholly moved mover" is O'Connor's (1995b: 174).

[60] Thanks to Tom Flint for causally influencing my freely thinking about this point.

claim, a claim the AC proponent rightly rejects. Since one's directly producing something intuitively is itself a doing, the AC theorist hears the purportedly jarring claim as saying that nothing and no one produces my directly producing something. But we should not obviously be jarred or troubled by this claim.

Moreover, the AC theorist invites us to reflect on the nature of a directly free action. Once we do, we'll see that it's perfectly licit to hold that nothing and no one causes one's directly free action—in the strong sense of 'cause'. Suppose, for sake of argument, that something produces an agent's directly free action. But if something produces one's action, there would be no principled way for meeting Whenceness with respect to that action. For, there are only three types of candidates for what might produce an agent's action: some event, someone else (i.e., another agent), or the agent herself. If someone else produces the agent's action, then there is no origination on the part of the agent and so the act is not free. If an event produces the agent's action, then there is no principled way to rule out deterministic causation and so, again, there is no origination on the part of the agent. If the agent's action can in principle be produced by something (the agent herself, e.g.), then this action could in principle be deterministically caused by some antecedent event, in which case the action does not strictly imply that the agent is an underived originator. For, if an event x can be produced, then x can be deterministically produced, *ceteris paribus*. Finally, and most importantly, we've seen good reasons for thinking that nothing could produce an ACE. So, for someone who thinks that one's directly free action just is an ACE, there are good reasons for thinking that nothing can produce one's directly free action.[61] Given AC, the counterintuitive claim is really the claim that nothing and no one produces one's directly producing something. But, again, this claim is not jarring. Rather, the jarring claim is that one produces nothing when one acts freely. Advocates of Goetz's view encounter this jarring claim head on, and the AC theorist may maintain that these advocates ought to be a bit troubled and jarred.

So, both Goetz and I think that reflection on the nature of a directly free action shows that nothing can produce it. We've both given arguments for this claim. I put it to my reader to compare our arguments. Which argument has the least controversial premises? Goetz simply asserts that whatever is produced is essentially an event with respect to which its agent is passive. So, nothing can produce any action, a fortiori nothing can produce a directly free action.

5.6   Conclusions and Work Left To Do

Acting freely intuitively requires that one is an ultimate source of change in the world. I have aimed in this chapter to express with precision this condition of origination that free will requires, calling the condition Whenceness. I discussed how various theories of free action appear to make provisions for meeting it. I investigated Ginet and Goetz's simple indeterminist theories, arguing that neither obviously meets Whenceness. I discussed Kane's theory of causal indeterminism. I showed that while Kane painstakingly attempts to meet the condition of origination, his attempt fails. I outlined the basics of O'Connor's theory of agent-causation, arguing that it obviously and straightforwardly satisfies Whenceness in a principled manner.

A crucial element of my case rests on the claim that nothing can cause an agent's directly free action—in the strong sense of 'cause'. For, on the agent-causal view, an agent's directly

---

[61] I also argue that the reasons for thinking that an agent produces her directly producing an event are unconvincing. See my assessment of Chisholm's case in Chapter 7, §7.4.

free action just is the agent's strictly and literally directly producing some event, and it is impossible to produce an agent's strictly and literally directly producing an event. I provided several arguments for this pivotal claim. I noted that the AC theorist's contention that nothing could produce an agent's directly free action confirms our pre-theoretic commitment to incompatibilism (Chapter 3, §3.3).

If the results of this chapter are correct, then there are substantial reasons for advocating the theory of agent-causation. Admittedly, though, these reasons are defeasible. For it may turn out that agent-causation's virtue in uniquely meeting Whenceness is defeated by its many vices or obscurities. Questions remain. For example, what more can be said about the nature of an agent-causal event (ACE)? How are agent-causation and event-causation related? For example, condition (iv) of the agent-causalist's four conditions of sole authorship (given above in §5.5) says that something may causally influence an agent's self-forming action (i.e., directly free action). But how is this supposed to work if nothing can produce an ACE? It seems plain that various factors can cause one's acting freely. So, how does the AC proponent resolve this apparent tension? The answer, I contend, lies in the claim that there may be mere causal contribution of an event in the absence of full-fledged causal production of that event. The distinction, here, is between causation in the strong sense of 'cause', on the one hand, and causation in the weak sense of 'cause', on the other hand. I gave this distinction an informative though incomplete treatment in §5.4. In the next chapter, Chapter 6, I work toward to unpacking this distinction by advancing a theory of causal contribution by events.

Chapter 7 then entertains many popular objections against agent-causation that fool many respected philosophers. I attempt to meet or disable them piecemeal, slowly earning the claim that the good reasons for adopting agent-causation (found in this chapter) are undefeated.

I postpone treatment of one objection until Chapter 8. Detractors claim that agent-causalists cannot provide sufficient and informative conditions for *rational* free action—that is, for an agent's acting freely *for* a reason. Obviously many free actions are rational. But how does agent-causation accommodate this truth? For example, if an agent's directly free action is not even produced by anything—not even her own reasons—then how does the agent act *for* or *on* her reasons? How do her reasons help explain her action without causing her action? How can an agent have multiple reasons some of which favor incompatible actions, and yet the agent act only on some of them? I address this purported Achilles' heel of agent-causation in Chapter 8. There, I outline O'Connor's attempt to answer these questions, and I defend his attempt from a putative counterexample. I argue that even if one concedes that these counterexamples stick, one could fix O'Connor's account by wedding it to my theory of causal contribution. My development of the agent-causal view in large part relies on being able to drive a wedge between strong causation and weak causation, i.e., between production and mere causal contribution. Let's turn, then, to seeing whether one can make sense of the general notion of mere causal contribution.

CHAPTER 6

MERE CAUSAL CONTRIBUTION

*One of the most common errors to which discussions of freedom and causation
are subject is that of confusing partial or contributing causes with
sufficient causal conditions.*
—Roderick Chisholm (1995: 97)

This chapter works toward explaining how agent-causation and event-causation are related. There is a *prima facie* distinction between the full-bodied causal relation of producing an event, on the one hand, and the more diluted and less potent causal relation of mere causal contribution, on the other hand. While I find the distinction fairly intuitive, I aim in this chapter to unpack the notion of causal contribution in considerable detail. This enables me to buttress O'Connor's agent-causal theory with a rigorous theory of causal contribution. The payoff is additional explanatory power and hopefully more credibility. For example, an account of causal contribution enables one to see from a theoretical perspective how something could contribute causally to an agent's directly free action even though nothing produces the agent's directly free action. That an agent-causalist can account for how agent's directly free action may have causal influences takes the sting out of a few objections (answered in Chapter 7) against our theory of agent-causation. For example, one such objection states that it is obviously false that nothing could cause an agent's directly free action, which contradicts your agent-causalist's commitments.

It is worth noting that every account of free action relies on the notion of an event's causally influencing a finite agent's free action. I leave it an open question whether or not an agent's reasons are among those things that *causally* influence her directly free action. Clearly reasons can influence free actions. Perhaps this influence is causal. Perhaps the influence is non-causal. It is just not obvious either way.[1] I shall take up in Chapter 8 the dispute over whether or not an agent-causalist can give a reasons explanation for a free action.

My point here is different. Even if reasons non-causally influence an agent's free action, that there could be other sorts of causal influences on an agent's free action seems quite intuitive. Thus, the results of this chapter aim to benefit everyone, as an adequate theory of causal contribution by events could be wedded to any good account of free action. This chapter, then, is a sort of icing on the cake. For one favoring cake over its icing, one could take as primitive the notion of causal influence and skip to Chapter 7 where I critically assess various popular objections against agent-causation.

Now my theory of causal contribution is not entirely of my own making. Instead, I carve out my own theory of causal contribution in the spirit and on the shoulders of Roderick

---

[1] Personal conversations with Tom Flint have helped me take this point more seriously.

Chisholm. I first consider Chisholm's (1995) final account but find it inadequate. I then turn to the best of his earlier accounts.[2] I entertain a sequence of unpublished objections each of which I either meet by mending the account's defects or undermine by showing how it misfires. The result is my account. I then show how one armed with my account could respond to a few published objections—advanced by William Rowe (1982) and Randolph Clarke (1996a)—against an earlier version of Chisholm's theory.[3]

Before delving into Chisholm's theories, let me first motivate in a bit more detail why one should take seriously the distinction between causal production and mere causal contribution. In the last chapter we noted that, according to the AC theorist, an agent's directly free action just is an agent-causal event (ACE) that is *the agent's causing some state of intention to act in a certain way*. We also noted that the concept of causation is one of our most fundamental ideas, that analyzing *the* causal relation with other more fundamental concepts is difficult. However, we noted that the AC proponent is a causal realist.[4] We intuitively take '*a* is *the* cause of *b*' to mean '*a* produces *b*' or '*a* generates *b*' or '*a* makes it that case that *b* occurs.'[5] Timothy O'Connor (2000a: 67), holds that

> agent causation is a distinct embodiment of the same primitive feature of causal production…at work in event causation, the difference consisting in the way that certain properties contribute to the causal potentialities of objects that have 'active power.'

Let '*a* is the *determinant* of *b*' abbreviate '*a* is *the* cause of *b*; *a* is that which *produces b*.' Now at the end of last chapter we examined in detail O'Connor's conjecture that an agent-causal event (ACE) cannot be produced by anything, that there is no determinant of a directly free action. Nonetheless, it seems highly unreasonable to hold that there cannot be at least some causally contributing factor on one's directly free action. Surely an ACE might be causally influenced by antecedent events, for any libertarian theory should permit some causal influence—however slight it may be—of one's directly free action. Clearly the complete circumstances in which finite agents freely act contribute causally to the agent's freely acting. Thus, the AC theorist should welcome the claim that there may be causal influences on an ACE.[6]

---

[2] Chisholm's theory of causal contribution was always in the workshop. The 1986-version I consider is far superior to his 1971-, 1979-, or even 1995-version.

[3] Even O'Connor rejects a version of Chisholm's theory of causal contribution. In Chapter 8, I consider and undermine O'Connor's (2002b: 349) objection.

[4] See §5.3. On *the* full-bodied causal relation, O'Connor (2002a: 136) reports:
I am a strong sort of realist about causation, holding that the basic causal propensities of the world are ontologically irreducible to purely 'qualitative' properties and the patterns of their distribution. … X's causing Y is not identical, in particular, to its raising the probability of Y, nor can it be given a more complicated analysis based on this notion.

[5] See O'Connor's (1995b: 175-7) brief exposition for taking the concept of causality as the "*primitive* notion of the 'production' or 'bringing about' of an effect."

[6] I use the terms "influence/influences" and "contribution/contributes" interchangeably. A tempting idea is that some x influences y only if x has an effect on y. But x has an effect on y only if x at least causally contributes to y. Nevertheless, perhaps there could be influence without causal

Taking these claims seriously seems to square with the pre-theoretic intuition that while nothing *produces* one's directly free action, there can be things that causally influence one's directly free action.

If the pre-theoretic intuition is right, it seems that there is room to make sense of an agent's being a mover while the agent is neither a wholly-moved-mover nor a wholly-*un*moved-mover. Intuitively, an agent is wholly moved only if there is something that makes her movement (i.e., her activity) occur. Intuitively, if there is something that counts as the determinant of the agent's acting a certain way, then the agent's movement/activity wholly moved by that cause. It is also intuitive that, since causal contributors "help move," an agent is wholly *un*moved precisely when there is nothing that even causally contributes to her activity. Perhaps God, in creating contingent entities *ex nihilo*, would count as a wholly unmoved mover.

The idea, then, is to try to make sense of the thought that a free agent is neither a wholly-moved-mover nor a wholly-unmoved-mover. In more positive terminology, when acting freely, an agent is a merely-partially-moved-mover. That is, the free agent is a mover whose movement is influenced by previous factors. However, the influence does not *produce* the movement. There are influences none of which count as the determinant of the agent's activity. The influence is merely…well…influential. It merely causally contributes.

We should be clear that *x's causally influencing y* does not obviously imply that x produces y. For example, I awake upon hearing the clock's alarm. The presence of oxygen in the room causally influences my waking up. This is confirmed by the fact that the most detailed explanation of my waking up right then and there will include the presence of oxygen in the room. So, *oxygen's being in the room* contributes causally to my waking up. *Oxygen's being in the room*, however, does not (even when coupled with the natural laws) make me wake up, although some collection of events presumably does.

Philosophers sometimes fail to appreciate this difference between a determinant and a mere-causal-contributor. Chisholm (1995: 97) writes, "One of the most common errors to which discussions of freedom and causation are subject is that of confusing partial or contributing causes with sufficient causal conditions." A sufficient causal condition is clearly a determinant. Philosophers often take too little care to discriminate between production and mere causal influence. They often make the issue one of either being caused or being uncaused. For example, Ted Honderich (1988: 389) says, "We have a pre-philosophical and pre-theoretical idea or whatever of a determinate centre, a self, which is uncaused in its activity and which is not superfluous."[7]

The property *being uncaused* is ambiguous. On the one hand, it could mean that there is nothing that counts as the determinant of event x. On the other hand, it might mean that nothing counts as a causal influence on x, i.e., that there is no causal contributor to x.

Or consider John Fischer's (1999: 107) claim that "the agent-causal event is not thought also to be subject to event causation." The property *being subject to event causation* shares the same ambiguity as the property *being uncaused*. What does *being subject to event* mean? Does

---

contribution. Perhaps God's reasons, e.g., influenced His actualizing a state of affairs in creating the world without contributing causally to His actualizing this state of affairs.

[7] See also Ginet (1997: 86), Chisholm (1966: 17), Wolf (1990: 14), Clarke (2003: 135), O'Connor (2002a: 136), and A.J. Ayer (1954: 275).

it mean that there is no event that is a determinant of an ACE? Or does it mean that there is no event that even causally contributes to an ACE?

Even Chisholm (1966: 11) once described the indeterministic view of human action as "the view that the act, or some event that is essential to the act, is not caused at all." *Event y's being not caused at all* strongly suggests that there is nothing that even contributes causally to y. But human action, whether free or not, is influenced by a host of factors. Thus, the property *being not caused at all* should be read as *having no determinant* rather than *having no causal contributor*.

In a sustained effort to bring rigor to this vital distinction, Chisholm advances several versions of an account of an event's contributing causally to another event.[8] Indeed, Chisholm's chief motivation for detailing a theory of causal contribution is his recognition that a satisfactory account of undetermined free action describes how various factors might causally influence without producing a freely performed action. Let's turn, now, to Chisholm.

6.1   On Roderick Chisholm's 1995-Account

Adequate discussions of a substantive Chisholmian theory invariably involve getting clear on a sequence of definitions. For better or for worse, the present discussion is no exception. Understanding Chisholm's characterization of causal contribution requires grasping his definition of an event.

Chisholm takes as primitive the fact that there is a substance having a property. For some substance x and property F there is a *state* x-exemplifying-the-property-F precisely when x is F. States have substrates and contents. Anything—a substance, a property, a state, etc.—exemplifying a property is called a *substrate*. Any property exemplified by a substrate is called a *content*. We see that there is an infinite hierarchy of states. A *first-order state* has a substance as its substrate, and a *second-order state* has a first-order state as its substrate. In general, an *(n+1)-order state* has an n-order state as its substrate, for any $n \geq 1$. For now, though, we need only be interested in first- and second-order states. So, a substance's exemplifying a property is called a *first-order state*, and a first-order state's exemplifying a property is called a *second-order state*.

According to Chisholm, events are a subspecies of states. Chisholm (1995: 96) defines an event as follows.[9]

(A)   x is an *event* =df. x is either a first-order state or a second-order state.

Intuitively, there are events that just are one event contributing causally to another event. So, some events have the property *being a contributing cause of event so-and-so*. Thus, there is motivation for including second-order states as events.

Understanding Chisholm's account of causal contribution also requires grasping his definition of a *minimal* sufficient causal condition, which in turn requires understanding his construal of a *sufficient causal condition*. Chisholm (1995: 97) defines the latter by reference to the properties designated as the contents of states or events:

---

[8] Chisholm (1967: 414-5 note 7; 1971: 45; 1976b: 201, 210 note 4; 1986: 59-64; 1995: 95-100).

[9] See also Jaegwon Kim (1976).

(B)   Set S of properties is a *sufficient causal condition* of property E =df. S is a set of
          properties such that the conjunction of its members does not logically imply E;[10] and it is
          a law of nature that, if all the members of S are exemplified by the same thing at the same
          time, then E will be exemplified either at that time or later.

Chisholm takes as primitive the laws of nature, which capture the sort of necessity that is
imposed by nature and not by logic alone. One might be tempted to think that (B)
straightforwardly yields in the following definition of causal contribution:

*   That state which is x-being-C *contributes causally* to that state which is y-being-E =df. C is a
    member of some set S of properties that are all exemplified by x at the same time, and S is a
    sufficient causal condition of E.

          But this will not do. For it implies that there could be cases of an intuitively causally
irrelevant property that is nonetheless causally relevant. Proof: Suppose C is a sufficient causal
condition of E. Consider some property P that is obviously superfluous to the etiology of
anything's exemplifying E. For example, consider the property, P, *being either green or not
green*. Now take the set C* of properties whose members include only P and the members of C.
If C is a sufficient causal condition of E, then so is C*. Intuitively, since P is by design causally
irrelevant to whether or not E occurs, then C* is no more causally relevant to the occurrence of E
than is C. So, contrary to the definition, P should not count as a causally relevant property.

          There are also counterexamples along these lines that do not rely on necessarily
exemplified properties like *being either green or not green*. We just need to find some
instantiated property that is obviously causally irrelevant to the exemplification of E, take the set
of properties whose members' exemplifications are included in the entire etiology of the
exemplification of E, and then add the irrelevant property to this set. For example, if 'E' denotes
being the winner of the World Cup, then *being a fly on the Wall of China* might be such a
causally irrelevant property.

          Seeing that one's definition of causal contribution should exclude causally irrelevant
factors, Chisholm (1995: 98) advances the notion of a *minimal* sufficient causal condition, which
states,

(C)   C is a *minimal* sufficient causal condition of E =df. C is a sufficient causal condition of
          E; and no subset of C is a sufficient causal condition of E.

Finally, Chisholm's (1995: 98) account for an event's contributing causally to another event goes
as follows:

(D)   That state which is x-being-C *contributes causally* to that state which is y-being-E =df. C
          is a member of some set S of properties that are all exemplified by x at the same time,
          and S is a minimal sufficient causal condition of E.

---

[10] S is a set of properties such that the conjunction of its members logically implies E *iff* it is
impossible that each member of S is exemplified but E is not exemplified.

6.2   How Chisholm's 1995-Account Fails

Chisholm's account falls prey to a few objections.[11] The first objection illuminates a slip on Chisholm's part when formulating (D). The problem lies in the fact that the definiens nowhere refers to the substrate of the event putatively having a causal contributor in Chisholm's definiendum. To see how this is a problem, consider the following example. In July of 1881, Charles Guiteau exemplified the property *having shot President Garfield*. Even though Garfield would probably have survived had it not been for the meddling of incompetent doctors, it is obvious that the event *Guiteau's shooting Garfield* contributed causally to the event *Garfield's being killed*. Now consider the bullet shot into Garfield as well as the property *being killed*. It was not ever the case that the bullet was killed. Guiteau's bullet was not the sort of creature that could be killed. Thus, it is impossible that the event *Guiteau's bullet's being killed* could occur.

Chisholm's account, however, has the result that *Guiteau's shooting President Garfield* contributed causally to *Guiteau's bullet's being killed*. Indeed, so long as there is a minimal sufficient causal condition (whose members are exemplified) of the property *being killed*, Chisholm's account implies that everything has this property at, or later than, the time it was exemplified as a result of the obtaining of the members of this minimal sufficient causal condition. Thus, the problem with Chisholm's account is that (D)'s definiens does not refer to the substrate of the event putatively having a causal contributor in the definiendum. The definiens should respect the fact that properties can be exemplified by more than one thing and should therefore exclude unwelcome participants.[12]

Fixing this difficulty is easy. We simply repair three of Chisholm's definitions as follows:

(E)   That state which is x-being-C *contributes causally* to that state which is y-being-E =df. C is a member of some set S of properties that are all exemplified by x at the same time, and S is a minimal sufficient causal condition of y-being-E.

(F)   Set of properties C is a *minimal* sufficient causal condition of the event y-being-E =df. C is a sufficient causal condition of y-being-E; and no subset of C is a sufficient causal condition of y-being-E.

(G)   Set S of properties is a *sufficient causal condition* of the event y-being-E =df. S is a set of properties such that the conjunction of its members does not logically imply y-being-E; and it is a law of nature that, if all the members of S are exemplified by the same thing at the same time, then y-being-E will occur either at that time or later.

There is one objection, though, that is not so easily circumvented. The objection may best be understood in light of considerations that help motivate Chisholm's account in the first place.

As a libertarian, Chisholm realizes that a free undertaking cannot have some sufficient causal condition whose members are exemplified. However, Chisholm attempts to soften the

---

[11] My objection also strikes against Chisholm's understanding of causal contribution in both (1971: 45) and (1976b: 201).

[12] When appropriate, I underline the point of adjustment between an inadequate definition and the proposed amended definition.

charge that such undertakings would be completely arbitrary and random by stressing that undertakings may have many causal contributors. Chisholm (1995: 99) claims, "…[T]here are…ways of contributing causally to an event that nevertheless has no sufficient causal condition."

Chisholm provides an example illustrating his claim that there are ways of contributing causally to an event that nevertheless has no sufficient causal condition whose members are exemplified. Someone shouts "Fire!" while you are in the middle of a room. You hear the scream, freely undertake to leave the room by the northern exit rather than by any of the others, and succeed. As Chisholm sets up the case, *your undertaking to leave the room* has a sufficient causal condition that is completed by your hearing the scream.[13] However, *your freely undertaking to leave by the northern exit* has no sufficient causal condition even though the shout contributed causally to it. Chisholm (1995: 99) states, "For even if the undertaking has no sufficient causal condition, there are several ways in which other events may contribute causally to that undertaking."[14]

Unfortunately, Chisholm's definitions imply that every event having a causal contributor also has a sufficient causal condition whose members are exemplified. Proof: Assume, for conditional proof, that x-being-C contributes causally to y-being-E. Together with (E) it follows that C is a member of a set S of properties that are all exemplified by x at the same time, where S is a minimal sufficient causal condition of y-being-E. Thus, there is a minimal sufficient causal condition of y-being-E whose members are exemplified. Since every minimal sufficient causal condition is a sufficient causal condition—as expressed by (F)—it follows that there is a sufficient causal condition of y-being-E whose members are exemplified. Therefore, completing our conditional proof, every event having a causal contributor has a sufficient causal condition whose members are exemplified. But this is intuitively false. Hence, Chisholm's account fails.

Moreover, one may say that according to the thesis of determinism, every occurring event has a sufficient causal condition all of whose members are exemplified. If this is a fair characterization of determinism, then Chisholm's account implies that, necessarily, every event has a causal contributor only if determinism holds. Many would agree that, necessarily, every event has a causal contributor. This does not bode well for Chisholm's account, for even Chisholm *qua* libertarian would agree that the thesis of determinism is not necessarily true.

6.3   A Superior Account
We've seen that Chisholm's 1995-account of event causal contribution is too restrictive. Since the thesis of determinism is possibly false, then possibly some events have causal contributors without having sufficient causal conditions. I propose that Chisholm's 1986-account of causal contribution is closer to the truth.[15] After advancing the account's five

---

[13] How does an event complete a set of properties? In a context where sufficient causal conditions are sets of properties, talk of a sufficient causal condition being *completed* and being completed by *an event* further complicates matters. This talk is a throwback to Chisholm's earlier theories of causal contribution, which I discuss below.

[14] Note that Chisholm's talk of an *event* having a sufficient causal condition confirms our adjustment of his definitions above.

[15] Chisholm (1986: 59-64).

fundamental definitions, I entertain a sequence of objections each of which I either meet by mending the account's defects or undermine by showing how it misfires.

6.3.1   Chisholm's 1986-Account

The reader may consult APPENDIX 1 for my finished account. Let "contributes" hereafter abbreviate "contributes causally." Here is Chisholm's (1986: 59-64) account of causal contribution, which I use as raw materials for my account.

(1)   Event C is a *sufficient causal condition* of event E =df. It is not logically necessary that if C occurs, then E occurs; and it is physically necessary that if C occurs, then E occurs either at that same time or later.[16]

(2)   Event C is a *necessary causal condition* of event E =df. It is not logically necessary that if E occurs then C occurs; and it is physically necessary that if E occurs, then C occurs either at that same time or earlier.

(3)   Property P is *conjunctive* =df. There are two further properties which are such that (a) neither implies the other, (b) P implies each of them, and (c) P is implied by whatever implies both of them.

(4)   Event C *completes* a sufficient causal condition of event E =df. There occurs a set of events such that (a) none implies a conjunctive property and (b) their conjunction is a sufficient causal condition of E; and C is a member of every such set.

(5)   Event C *contributes causally* to event E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E or (b) E occurs and D completes a sufficient causal condition of a necessary causal condition of E.

Let's assume the same definition of an event. Let 'SCC' abbreviate 'sufficient causal condition' and let 'NCC' abbreviate 'necessary causal condition'. Nearly all of these definitions are straightforward. Seeing what motivates (4) and (5), though, is not immediately obvious. A bit of reflection reveals that (4) captures Chisholm's attempt to exclude causally irrelevant factors from being deemed contributors. Suppose, e.g., that there is a SCC for *the house's catching fire*. There are other SCCs besides this one—indeed, some that properly include this one. Many of them contain superfluous "parts of events" such as *the bird's sitting on the roof*. But *the bird's sitting on the roof* does not contribute to *the house's catching fire* simply by being part of any SCC of *the house's catching fire*; rather, it needs to be part of a minimal SCC. Definitions (4) and (5) respect this, as a minimal SCC just is a completed SCC.

Unlike his 1995-account, this account leaves it open for an event having no SCC to nonetheless have a contributor. Perhaps the effected event nomologically requires another event,

---

[16] Let "physically necessary" abbreviate "it is a law of nature". Notice that C's not occurring is consistent with C's being a sufficient causal condition for some other state. More precisely, C, which is some first- or second-order state x-being-F, might be a sufficient causal condition for some other state without C's obtaining.

which may properly be said to enable the effected event. For example, while *your freely undertaking to leave the room by the northern exit* has no SCC, it nomologically requires some or other events. Perhaps *person X's screaming "Fire!"* completes (or, more exactly, implies an event that completes) a SCC of one of those nomologically required events. Thus, *person X's screaming "Fire!"* contributes to *your freely undertaking to leave by the northern exit*.

6.3.2   Revising Chisholm's Account in Light of Unpublished Objections

Now for the unpublished objections, which are interesting for their own sake. The Redundancy Objection: Definitions (4) and (5) exclude more than just causally irrelevant factors. They exclude relevant factors as well. Proof: Think about cases of redundant causation. For example, the following seems possible. C is a member of a set of occurring events none of which implies a conjunctive property but that together they count as a SCC of E, C* is a member of a distinct set of occurring events none of which implies a conjunctive property but are such that they together also count as a SCC of E. By hypothesis, neither C nor C* is a member of *every* set that minimally counts as a SCC of E. If we further suppose that C does not complete a SCC of a NCC of E (which seems consistent with the above assumptions), then the account implies that C does not contribute to E, which is absurd. Surely, such a case of redundant causation is possible.[17]

Solution: Broaden the definition of *completing* a SCC to permit cases of distinct SCCs of a single event while still excluding cases of intuitively causally irrelevant factors that the original definition was designed to handle.

- Event C *completes* a sufficient causal condition of event E =df. There occurs a set S of events such that (a) none implies a conjunctive property, (b) their conjunction is a sufficient causal condition of E, ~~and C is a member of every such set;~~ and (c) no conjunction of any proper subset S* of S is a sufficient causal condition of E; and C is a member of *some* such set S.

The-All-Necessary-Causal-Conditions-are-Contributors Objection: The account excludes intuitive contributors. Proof: Consider a free undertaking E. There is no SCC of E, but there are NCCs. Intuitively, *every* NCC of E should count as a contributor of E. Consider some event F

---

[17] A similar problem arises even in cases that do not involve redundant causation but rather involve two distinct SCCs of E occurring in temporal sequence. Suppose C completes a SCC, F, of E. Thus, there occurs a set S of events such that (a) none implies a conjunctive property and (b) their conjunction—i.e., F—is a SCC of E; and C is a member of every such set S. Let no member of S occur earlier than time T. Suppose also that C* completes a SCC, F*, of F. Thus, there occurs a set S* of events such that (a) none implies a conjunctive property and (b) their conjunction, F*, is a SCC of F; and C* is a member of every such set S*. Suppose now that every member of S* occurs *before* T. Since C* occurs before T and since no member of S occurs before T, then C* is not a member of S. Thus, C* is not a member of every SCC of E. Thus, C* does not complete a SCC of E. However, the relation of being a SCC is surely transitive. Thus, since F* is a SCC of F, and F is a SCC of E, then F* is a SCC of E. Since C* completes F* and F* is a SCC of E, it follows that C* completes a SCC of E. So, C* both does and does not complete a SCC of E, which is absurd. Therefore, Chisholm's definition of *completing* a SCC is too restrictive.

144

that is a direct NCC of E—i.e., F is a NCC of E but is not a NCC of a NCC of E. Thus, F occurs either at the same time as E or at a time immediately preceding E. By Chisholm's lights, F is a contributor of E only if F implies an event that completes another NCC of E. But this just seems wrong. F seems to be a contributor of E simply by being itself a NCC of E, not by going through some other NCC of E. In sum, it seems that every NCC of an effected event should count as a contributor.

Solution: Concede that every NCC contributes to the effected event. So…

- Event C *contributes causally* to event E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E, or (b) E occurs and D <u>either is a necessary causal or</u> completes a sufficient causal condition of a necessary causal condition of E.

The Unspecific-Events-Have-Contributors Objection: The account is too restrictive, ruling out intuitive contributors. Argument: Let E be *Garfield's dying*, and suppose this event, perhaps due to some inherent random feature, has no occurring SCC. Even though *Charles Guiteau's shooting Garfield* contributes to E, there are possible worlds where the laws L hold, Garfield dies, yet there is no event D being such that D either just is or is implied by *Guiteau's shooting Garfield* and D completes a NCC of *Garfield's dying*. For, there is a world governed by L where Garfield dies as an infant. Intuitively, in some infant-Garfield-world there is no event D implied by the state *Guiteau's shooting Garfield*, where D either is a NCC or completes a SCC of a NCC of *Garfield's dying* in this infant-Garfield-world.

Solution: There are other, "more expansive" events (e.g., *Garfield's dying in circumstances so-and-so*, or perhaps *Garfield's dying a death of type such-and-such*) that imply and are contemporaneous with *Garfield's dying* that have contributors according to the above account.

- Event C *contributes causally* to event E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E, or (b) E occurs and D either is a necessary causal condition or completes a sufficient causal condition of a necessary causal condition of <u>some event E* that implies and is contemporaneous with</u> E.

The Superfluous-Contemporaneous-Events Objection: The account is too broad, permitting too many contributors of E. Proof: Suppose that C contributes to some event E in virtue of implying an event D that completes a SCC of a NCC of some event E* that implies and is contemporaneous with E. Now consider an event F that occurs simultaneously with E but where <u>C does not contribute to F</u>. Let F* be the conjunctive event *its coming to pass that E* and F occur*.[18] Any NCC of E* is a NCC of F*. Thus, C implies D, D completes a SCC of a NCC of

---

[18] This assumes a sort of mereological essentialism with respect to events that occur simultaneously. For any two simultaneously occurring events, there is an event that is their conjunction. One might inquire about whether or not there could be conjunctive events. Maybe there are only two events here: *its coming to pass that E** and *its coming to pass that F*. Nonetheless, if both E* and F occur, then there is a conjunctive property, P, *being such that E* and F occur*. Moreover, P is obviously exemplified by some substance *x*. Hence, by Chisholm's definition of an event—expressed in (A)—it

F*, and F* implies and is contemporaneous with F. Thus, the account implies that <u>C contributes to F</u>, which contradicts our hypothesis. Thus, the account is too broad, incorrectly labeling C a contributor of F.

   Solution: The objection relies on the fact that F* has a part (*viz.*, F) that is causally superfluous to the occurrence of E. Since E* logically implies E, so does any event larger than E*. But there are many events larger than but contemporaneous with E* that shouldn't figure into the occurrence of E, and F* is obviously one of them. By design, F* has a part that is causally superfluous with respect to E, which clearly resembles a problem we addressed above. We therefore employ an analogous resolution.

- Event D *enables* E *by a minimal canopy condition* E* =df. E occurs; <u>there is a set S of events such that (i) none of S's members implies a conjunctive property and the conjunction of the members of S is equivalent to event E*;</u> (ii) E* implies and is contemporaneous with E; (iii) D either is a necessary causal condition or completes a sufficient causal condition of a necessary causal condition of E*; <u>and (iv) D neither is a necessary causal condition nor completes a sufficient causal condition of a necessary causal condition of a proper subset of S</u>.

- Event C *contributes causally* to event E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E, or (b) <u>D enables E by a minimal canopy condition E*</u>.

   The Knowledge-Contributes-Too-Much Objection: The account is too lenient, permitting too many contributors. Proof: Let C be *Sam's knowing that D occurred 100 years ago*. Then, if D contributes to some event, then so does C; hence, we have backward causal contribution, which is absurd. A tempting solution is to build a rejection of backward causation into the definition as follows.

- Event C *contributes causally* to event E =df. C occurs <u>but not later than E</u>; and C implies an event D <u>such that D does not occur earlier than C and</u> either (a) D completes a sufficient causal condition of E, or (b) D enables E by a minimal canopy condition E*.

However, the problem is not only with *when* the event *Subject S's having the true belief that D occurs* occurs. The problem is more general. Suppose it is clear that *Smith's ingesting cyanide*[19] completes some SCC of either *Smith's dying* or some NCC of *Smith's dying*. Suppose also that at the very same moment Smith ingests cyanide, Cuthbert comes to believe truly that Smith ingests cyanide. From these assumptions, Chisholm's account implies that *Cuthbert's having the true belief that Smith ingests cyanide* contributes to *Smith's dying*. However, clearly it is possible that these assumptions are consistent with it not being the case that *Cuthbert's having the true belief that Smith ingests cyanide* contributes to *Smith's dying*.

---

follows that there is some event *x's-exemplifying-P*, which does the same duty as what we were calling the conjunctive event. Thanks to Tom Flint for bringing this worry to my attention.

  [19] Or, if this event implies a conjunctive property, replace it with an appropriate substitute.

This sort of objection also strikes against the previous introduction of E* into the definition. E.g., suppose E* is *Cuthbert's having the true belief that E occurs*, where intuitively C contributes to E* but not to E. Again, our account wrongly implies that C contributes to E.

Solution: Our solution rests on two obvious points. First, the events presenting problems for the account are complicated critters in that they imply conjunctive properties. More basic or atomic events do not present the same sort of difficulty. But we don't want to preclude non-atomic events from being contributors. Second, the causal contribution relation is obviously transitive. We realize that, necessarily, if A contributes to B, and if B contributes to C, then A contributes to C.

With this in mind, the objection seems to require that some contributors of *Cuthbert's having the true belief that Smith ingests cyanide* are not contributors of *Smith's dying*. Part of our solution, then, is to learn from the fact that C contributes to E only if every contributor of C is a contributor of E. Our account so far is superior to Chisholm's but still not without residual defects (as illustrated by the objection under consideration). The core of our account, I think, seems promising. So, let's define a *global causal condition* according to our account presently under scrutiny and then adjust our definition of causal contribution to include the clause that every global causal condition of a global causal condition C of E is a global causal condition of E. Moreover, to avoid the complicated critters from unnecessarily complicating matters, we may restrict the global causal conditions of the putative contributor C to those not implying a conjunctive property. No additional primitive notions are introduced.

- C is a *global causal condition* of E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E, or (b) D enables E by a minimal canopy condition E*.

- Event C *contributes causally* to event E =df. C is a global causal condition of E; and every atomic global causal condition F of C is a global causal condition of E.[20]

Notice that this solution meets the recently cited difficulty against introducing E* into the definition. To illustrate the difficulty, suppose E* is *Cuthbert's having the true belief that E occurs*, where C contributes to E* but not to E. The critic thinks that the account wrongly implies that C contributes to E. To the contrary, now that we've revised the account, it no longer obviously implies that C contributes to E. For, there may be atomic global causal conditions of C that are not global causal conditions of E.[21]

---

[20] Let an *atomic* global causal condition be a global causal condition not implying a conjunctive property.

[21] For instance, consider the following assignments, F = *Vince's seeing Cuthbert*, C = *Vince's telling Jones that Silver Charm won the Derby*, E* = *Cuthbert's having the true belief that Silver Charm won the Derby*, and E = *Silver Charm having won the Derby*. Intuitively, C contributes to E* but not to E. Moreover, F is as good a candidate as any for an atomic global causal condition of C. However, it is not the case that *Vince's seeing Cuthbert* is a global causal condition of *Silver Charm having won the Derby*. If F is not atomic, consider the set of occurring atomic events that are together equivalent to F and then consider whether each of these is a global causal condition of E). And even if F happens to somehow be a global causal condition of E, there would seem to be some other atomic global causal condition of C that is not a global causal condition of E.

Working the notion of a global causal condition into our account simultaneously preempts an independent worry, indicating that our recent alteration is motivated and principled. The worry suggests that our theory is too broad. To illustrate, suppose that *the frost being late* contributes to *the apple harvest being poor*. The following year there is another poor apple harvest. Does the late frost contribute to the second occurrence of the poor apple harvest? Possibly not. And without our recent resolution, it would seem that we're unfortunately committed to thinking that the late frost contributes to *every* subsequent occurrence of the poor apple harvest—hence the suggestion that our theory is too broad. However, our new account implies that (i) the late frost contributes to the second occurrence of the poor apple harvest only if (ii) every atomic global causal condition of the late frost is a global causal condition of the second occurrence of the poor apple harvest, which need not be the case. Moreover, if it happens to turn out the (ii) holds, then it would appear that the late frost does in fact contribute to the second occurrence of the poor apple yield.

The Lack-of-Unity Objection: Your account is overly disjunctive, which has the whiff of *ad hoc* maneuvering. In particular, I'm thinking of your definitions of a global causal condition and of enabling an event by a minimal canopy condition, each of which include two disjuncts.

Reply: Yes, the conditions are in places expressed by disjunctive sentences or formulae. But I see no good reason to think that there are disjunctive *propositions*. These disjunctive sentences pick out the proposition that counts as an analysis for when one event contributes to another event. Moreover, our adjusted account uses no more primitive notions than those already found in Chisholm's original account.[22] As for the charge of ad hocery, I'm not defending Chisholm's account. I'm constructing my own. What you see here is part of the construction process.[23]

The Fallacy-of-the-Facile-Amendment Objection: Tom Flint tells me that James Ross coined a locution that may very well apply to my account.[24] The locution incites suspicion.

Allegedly, my account commits "the fallacy of the facile amendment." The idea, so I am told, is that I've started with a certain theory, encountered various objections, and simply tacked on a clause to parry each objection along the way. Then, after too many epicycles, I encounter

---

[22] The Transitivity Objection: Any adequate theory should get the intuitive cases right, and it's not obvious how your account handles the transitivity of causal contribution. Consider a case where (i) E has no occurring sufficient causal condition, (ii) C intuitively contributes to E in a derivative manner, as (iii) C occurs in E's very deep past. For example, if Cleopatra hadn't introduced Caesar to the Egyptian astronomer Sosigenes, Caesar would not have instituted the solar calendar as we know it today, leap year included. See Beyer (2003: x-1). Intuitively, *Cleopatra's introducing Caesar to Sosigenes* contributed in a round-about way to many recent events, say, *Jimmy's freely giving to OXFAM on a date designated by "February 29th"*. Shouldn't we just see how your account gets this case right? Reply: For those who find this objection convincing, we may without introducing any new primitives build the notion of transitivity into the definition as follows:

- Event C *directly* contributes causally to event E =df. C is a global causal condition of E; and every atomic global causal condition of C is a global causal condition of E.
- Event C *contributes causally* to event E =df. C is a member of some ordered n-tuple $\langle C_n,....,C_1 \rangle$ for any $n \geq 1$ such that both (i) C_1 directly contributes causally to E and (ii) for every pair of members $C_{(n+1)}$ and C_n, $C_{(n+1)}$ directly contributes causally to C_n.

[23] See also the end of APPENDIX 1 for an attempt at removing a disjunct.

[24] Personal conversation.

no more objections. Now, this may be due to the fact that I have the right account. Or, it may be because things have gotten so complex that no one wants to spend enough time to come to grips with it. So, good judgment suggests that one should be suspicious of the final result. Indeed, maybe I started off with the wrong theory at the outset.

Reply: I'm not sure how exactly to understand this objection. For at least two reasons, I am unclear why my objector would use the locution "the fallacy of the facile amendment." First, fallacies are fallacious inferences. Now, I make many inferences in the above construction of my account. I think that none of them are fallacious. But if one contends that I make a fallacious inference, then one should provide a counterexample to the inference in question. Second, a great deal of effort went into constructing these amendments. So, I fail to see how my amendments are facile.

I've thought long and hard about the conditions, and they seem right to me. Frequently, in my vigorous attempt to construct counterexamples against my own account, I learned that what I initially suspected was a counterexample was already handled by previous amendments. I think this lends credibility to the claim that these previous amendments were indeed principled and a natural way to develop the account. It's not as though I've simply deployed my complete account without explaining what work each condition or clause does. Rather, I outlined the journey I took in constructing an account.

But maybe there's something else to the fallacy-of-the-facile-amendment objection. Perhaps my objector assumes that if it's too hard for trained philosophers—better yet, well-trained philosophers, philosophers in the know, philosophers finely tuned to fine conceptual distinctions, philosophers equipped with razor sharp intellects, philosophers who perhaps even dream in S5—to figure out what it would take to construct a counterexample to the account, then one should reject the account. After all, if the complete account is too hard to understand, what good is it? Fortunately, my account isn't *that* abstruse. For example, Alvin Plantinga says of my account:

> I followed much or most of it. And it seemed to me to be right. I can't say that I got a full grasp of your complete conditions. I thought that was a bit much to ask of one sole philosopher like me to understand. It looks more like a team effort. Maybe Peter [van Inwagen] and I could figure it out together. … It might be absolutely right and it might be that there isn't anything simpler that's absolutely right. But even so, it's going to suffer from the fact that people aren't going to invest the effort to figure it out.[25]

Philosophy is hard work. There are compromises and trade-offs. The tension between maximizing a theory's rigor and maximizing its accessibility is well known. As Saki notes, "A little inaccuracy save tons of explanations."[26]

I wish I were able to make it simpler, elegant, and even beautiful. But the sheer fun of chisholming as well as the recognition of encroaching deadlines seemed to have determined that

---

[25] Personal conversation.

[26] Quoted in Mick Farren (2004: 102).

I do otherwise. I find a whiff of consolation in Nadine Gordimer's remark that "the truth isn't beautiful, but the hunger for it is."[27]

6.3.3 Meeting Published Objections

Two philosophers, William Rowe and Randolph Clarke, register complaints against a version of Chisholm's account of causal contribution that predates his 1986-account. While ignoring the question as to whether they undermine their intended target, seeing how they fail to undermine our present account is instructive.

Clarke's Objection. Clarke (1996a: 35) claims that, possibly, one create a SCC for an agent's movement and that this created condition has *no* causal influence on the agent's moving because the agent moves on the basis of her previously made decision to move. Even though the created SCC restricts the agent by removing his option to remain still, Clarke thinks it might not have any causal effect on the agent's movement.

On the contrary, it is no less reasonable to contend that *both* the agent's prior decision to move and the created SCC contribute to the movement. Our account permits multiple SCCs.[28] Recall that the created SCC does not logically imply that the movement occurs, and it is a law of nature that if the created condition occurs, then the movement occurs either at the same time or later. One may reasonably think this is a case of causal influence. The fact that the agent already decided to move and acts on this decision merely throws another contributor into the mix. The created SCC does not preclude the contribution of the agent's previous decision to move.[29] Likewise, neither does the contribution of the agent's previous decision to move preclude the contribution of the created SCC. Clarke's objection, therefore, fails to count as a *clear* counterexample. Of course, it goes without saying that the created SCC need not have any influence on the agent's previous decision to move in order to influence the movement itself. Backward causal contribution is impossible.

Rowe's Objection. Rowe's (1982) version of this objection can be put in the form of a challenge. He presents two examples. The first example illustrates an intuitive case of causal contribution. The second consists in altering the first example slightly, illustrating how the contribution is intuitively absent. The challenge is to honor these intuitions, and Rowe argues that Chisholm's account does not meet this challenge. Our new account, however, meets the challenge straightforwardly.

In the first example, we're to imagine that a fire completes a SCC for a man's jumping off a pedestal either to the north, east, south, or west. We suppose that the man has some reason to jump west. Indeed, given that he has to jump he would jump west, other things being equal, because west is the direction he wants to travel. Now just prior to jumping, he notices that the terrain to the west is unfavorable, which completes a SCC for *his not jumping west*. He then freely jumps to the east. Intuitively, *his noticing the unfavorable terrain* contributes to *his freely jumping east*. An adequate account should be consistent with—and perhaps even illuminate—this result.

---

[27] Quoted in Mick Farren (2004: 107).

[28] See the definition of *completing* a sufficient causal condition, APPENDIX 1.

[29] Moreover, there is a *tu quoque* objection to Clarke's criticism, as Clarke's (1996a) theory of agency requires that a single event have distinct causes each of which produces the event.

Our account easily deals with this case. For, it may be the case that *his noticing the unfavorable terrain* contributes (according to our account) to an event such as *his loosing a reason to jump west*—perhaps the former event implies an event that completes a SCC of *his loosing a reason to jump west*. Moreover, *his loosing a reason to jump west* may imply an event that enables *his freely jumping east* by a minimal canopy condition E*. Thus, according to our account *his loosing a reason to jump west* would contribute to *his freely jumping east* via contributing to *his loosing a reason to jump west*.

Rowe's second example consists in slightly altering the case above. The man does not generally prefer to jump to the west. Rather, the man generally prefers to jump toward the sun and it happens to be morning. Indeed, Rowe stipulates that he would freely jump east whether or not he notices the unfavorable terrain to the west. Even if *his noticing the unfavorable terrain* implies an event that completes a SCC for *his not jumping west*, intuitively the former does not contribute to the event *his freely jumping east*. Rowe (1982: 375, *my emphasis*) states,

> Although his noticing the terrain restricts his options, renders him unable to jump west, it does not make a causal contribution to his free act of jump east *since* he in fact would not have jumped west had he not noticed the terrain to the west, he would still have jumped to the east.

Again, our account is consistent with this intuition. First, there may be an atomic global causal condition of *his noticing the unfavorable terrain* that is not a global causal condition of *his freely jumping east*. Second, *his noticing the unfavorable terrain* may not itself be a global causal condition of *his freely jumping east*. For, since there is no SCC for *his freely jumping east*, then *his noticing the unfavorable terrain* implies no event that completes a SCC of *his freely jumping east*. Moreover, it is not clear that *his noticing the unfavorable terrain* implies an event that enables *his freely jumping east* by some minimal canopy condition. There does not seem to be an event implied by *his noticing the unfavorable terrain* that either (i) just is a NCC of *his freely jumping east*, or (ii) completes a SCC of a NCC E* that implies and is contemporaneous with *his freely jumping east*.[30] Thus, this example is not a clear counterexample to our account.

6.4   Conclusion

Every libertarian account relies on the notion of an event's merely-causally-contributing to a finite agent's directly free action. An adequate theory of causal contribution by events, then, should be wedded to any libertarian account. This chapter works toward making libertarian accounts more rigorous by finding such an adequate theory of causal contribution. I focused on Chisholm's ongoing attempts to do the same. I argued that his 1995-, 1986-, 1979-, and 1971-accounts of causal contribution fail. Finding his 1986-account most promising, however, I modified his account in order to handle a series of unpublished objections. We then saw that the finished product has the resources to handle some published objections as well. As we shall see

---

[30] This is not to say that there seems to be no event implied by *his noticing the unfavorable terrain* that either (i) just is a NCC of *his freely jumping east*, or (ii) completes a SCC of a NCC E* that implies and is contemporaneous with *his freely jumping east*. It does not *seem* to be the case either way. It's just unclear. But counterexamples should be clear.

in the next two chapters, the proponent of agent-causation may exploit the concept of causal contribution to explain her view more thoroughly as well as to turn away objections advanced against her view of agent-causation.

CHAPTER 7

DEFENDING AGENT-CAUSATION

The concept of agent-causation has a long history in philosophy,
its demise being a relatively recent affair.
—William Rowe (1991b: 238)

The agent-causalist's characterization of an agent's directly free action survives a host of objections, objections that many philosophers consider decisively damaging. Seeing how our view handles reasoned criticism aids in understanding the view itself. In the remainder of my book, I attempt to earn the claim that our AC view is defensible.

To this end, *this chapter* provides a reasoned agent-causalist response to the most popular objections against AC, with the exception of one. Arguably the most impressive or formidable objection states that the AC view lacks resources for explaining how an agent may freely perform an action *for a reason*. I put off addressing this objection until Chapter 8. Hence, if our arguments there are correct, an agent-causalist can render intelligible how an agent's directly free action may be performed rationally. As for what remains in this chapter, I briefly review our agent-causalist stance on what are the essentials in characterizing the nature of an agent's directly free action. I then consider many objections, concluding that none of them is clearly damaging.

7.1   Agent-Causation 201

Chapters 5 and 6 work toward characterizing several aspects of the nature of an agent's directly free activity.[31] Let's review some essentials.

First, the signature feature of our AC view is that an agent performs a directly free action only if an agent-causal event (ACE) occurs, i.e., only if it comes to pass that the agent *qua* substance strictly and literally directly produces an event. If there is an ACE, then Eventism is false, where 'Eventism' designates the thesis that, strictly speaking, only events produce events. Thus the agent-causalist emphasizes her claim that when an ACE occurs, there is no event-mechanism solely by which, through which, with which, or in virtue of which the agent directly produces an event. To demand such an event-mechanism would be just to deny the possibility of our AC view by denying its signature feature. An ACE cannot without loss of meaning be paraphrased into "event-speak." Finally, an ACE just is the agent-causalist's candidate for an agent's directly free action. We follow O'Connor in holding that the entire event-constituent of an ACE is a state of intention to act in a certain way. An ACE may be regarded as a decision.

Second, the AC theorist points to the ACE as the event satisfying Whenceness, thereby meeting the origination condition that free action requires. The agent-causalist minces no words in answering the question "Whence come the event-constituents of an agent's directly free

---

[31] And to answer the Intelligibility Question, the AC theorist provides sufficient conditions for the nature of an agent's directly free activity—conditions that essentially imply an ACE.

153

behavior?" The answer is, in the most literal sense possible, *the agent*. The agent is an obvious source, a not wholly derived source, of an element

intrinsic to her directly free action. Indeed, the agent is an ultimate originator of the entire event-constituent of her directly free action. The AC theorist, then, shares with the causal indeterminist the belief that the notion of origination is a causal phenomenon that therefore admits of some causal elucidation. We AC theorists do more than merely stipulate that the agent, when acting freely, is an underived originator.[32] We do more than assert that an agent originates an element essential to her directly free action. Moreover, we have seen good reasons to think that the leading Eventist accounts of free action fail to capture the origination condition in a principled way. Consequently, while we have too little space to investigate and refute every Eventist account, we affirm John Bishop's (1986: 228) observation:

> [A]ffirming that a concept is primitive is not equivalent to appealing to mystery, since such concepts may be illuminated by holistic rather than reductive means (charting their criteria of applicability, exposing the structure of the conceptual network to which they belong, etc.). Indeed, if the primitivism of the agent-causalist is accompanied by a convincing case for the inadequacy of the best going event-causalist reductions…, Agent-Causalism may be judged well confirmed.

Third, we concluded in Chapter 5 that nothing can causally produce an agent's directly free action. For, Eventism permits no principled way to account for the agent's being an underived source of some element implied by her directly free action. Moreover, this result independently confirms the possibility of libertarianism. For, if directly free action requires an ACE and if nothing can produce an ACE, then there can be no sufficient causal condition of an ACE. Hence, a directly free action precludes determinism.

Third, there may be events that causally contribute to an ACE in the manner prescribed by our account of event causal contribution outlined in Chapter 6.[33] Events obviously might causally influence an agent's directly free action. For example, I know that Mr. Crisp enjoys tying knots. Suppose that I encounter him busily drafting an essay expositing the relationship between causation and the philosophical thesis of Presentism. I interrupt him cordially and hand him a rope. Smiling, he briefly deliberates about which of the many complicated knots he'll attempt first. Suppose that he freely decides to tie a figure-eight knot. It should seem about as obvious as anything that *my handing Mr. Crisp a rope* contributes causally to *his freely deciding to tie a figure-eight knot*. However, *my handing Mr. Crisp a rope* does not produce *his deciding to tie a figure-eight knot*. In keeping with AC, we may suppose that Mr. Crisp's deliberation ends precisely when he, the man, strictly and literally directly produces a state of intention to tie a figure-eight knot. His directly producing this state of intention is a decision, and the AC

---

[32] Pacé Ted Honderich's claim that no expert on free will has anything informative to say about the relationship between an agent *qua* originator, on the one hand, and the decision that lies at the core of her free action, on the other hand. Honderich (1993: 42-3) states,

> To my mind the best response to the demand for an explanation of the relation between an originator and decisions is that an explanation cannot be given. We have to regard this relation as primitive or unanalysable. …[N]othing can be said about the relation itself. We just understand it.

[33] Chisholm (1978: 628-30) thinks that there could be events that contribute causally to an ACE.

theorist contends that it is a free decision. Intuitively, *my handing Mr. Crisp a rope* contributes causally to Mr. Crisp's ACE.

So, in general, causal influence need not be full-fledged causal production. While our view precludes something's causally producing an ACE, we welcome the claim that something could causally contribute to an ACE. And now that we have a theory of causal contribution, we can understand how an event might causally influence an agent's directly free activity. Grafting our theory of causal contribution onto O'Connor's AC view considerably deepens the view by noting how agent-causation and event-causation are related. O'Connor (2000b: 113) gestures at how they are related in the following passage:

> Freedom of the will, in my judgment, involves the exercise of a distinctively personal form of causality, one which differs in certain respects from the mechanistic form of causation operative in impersonal causal forces. In the mechanistic case, objects have specific causal powers, or dispositional tendencies, associated with their fundamental intrinsic properties. The powers might concern a unique outcome or range of possible effects that is structured by a specific probability measure. Either way, they exercise certain of these causal powers as a matter of course when they are placed in the appropriate circumstances. Such circumstances either stimulate a latent mechanism or remove inhibitors to the activity of a mechanism already in a state of readiness. Strictly speaking, the cause here is the *event* of the object's having these power-conferring properties in those circumstances.
>
> According to some of us, there is another species of the causal genus, involving the characteristic activity of purposive free agents. Such agents can represent possible courses of action to themselves and have desires and beliefs concerning those alternatives. Against that background motivational framework, they themselves directly bring about immediately *executive* states of intention to act in various ways. This direct causing by agents of states of intention goes like this: As with mechanistic causes, the distinctive capacities of agent causes ('active powers') are grounded in a property or set of properties. So any agent having the relevant internal properties will *have it directly within his power to* cause any of a range of states of intention delimited by internal and external circumstances. However, these properties function differently in the associated causal process. Instead of being associated with direct causal functions from circumstances to effects, they (in conjunction with appropriate circumstances) *make [causally] possible* the agent's producing an effect. These choice-enabling properties ground a different type of causal power or capacity—one that in suitable circumstances is freely exercised by the agent himself.

So I give a theory of causal contribution to minimize the charge that the agent-causalist cannot explain how the agent-causal relation relates to the event-causal relation. We would have reason to abandon our theory of agent-causation, if we learned that nothing could causally contribute to an ACE. Why? Because it's just obvious that something might causally contribute to an agent's directly free action, and, according to our AC theory, an ACE just is an agent's directly free action. Nonetheless, my theory of causal contribution is consistent with an event's

contributing causally to an ACE. So, no one should reject our theory of AC on the grounds that the AC theorist has no resources to explain how an event could causally influence an ACE.

To conclude our third point, even if our account of causal contribution is flawed, the AC theorist may hold that an agent's directly free action can have causal contributors, for this is not a thesis special to agent-causation. I should think that nearly everyone should believe that there could be causal influences on one's directly free action. Any complete and philosophically adequate theory of free action should have a theory of causal contribution. Thus, even if our account of causal contribution in Chapter 6 fails, then until there is an adequate account of causal contribution, the AC theorist, just like other experts of free action, may tentatively take as primitive the notion of causal influence.

Now that we have a fairly good idea about what some reasonable agent-causalists believe, let's see whether there is any good reason for thinking that we agent-causalists should give up our view.

7.2   Simplicity

Recall that 'Eventism' designates the thesis that strictly speaking only events produce events. Eventism, then, precludes agent-causation. John Bishop (1986: 229-30) states:

> [T]here is an obvious methodological reason for preferring Event-Causalism [or Eventism], namely the commitment not to multiply conceptual primitives beyond necessity. Agent-Causalism may be embraced only as the theory of second choice, if we find reason [for thinking that it is] impossible to satisfy the constraints of Event-Causalism [or Eventism].

Bishop suggests a popular principle. To illustrate the principle, suppose that two theories (A and B) compete in accounting for a given range of intuitive data. Suppose also that they account for this data equally well. However, suppose that the number of theory A's conceptual primitives is less than the number of theory B's. All other things being equal, preferring theory A to theory B would seem most reasonable. Most philosophers assume this as a central assumption of doing metaphysics properly. For, the additional posits of theory B do not explain or account for anything not already accounted for by the fewer posits of theory A. Thus the additional posits of theory B should be deemed as merely putative posits and so eliminated from any adequate theory. So much the worse for a theory having merely putative posits.

Even if the above characterization of doing metaphysics properly is correct, there is no damaging argument against AC here. For we saw in Chapter 5 strong reasons for thinking that Eventism does not in a principled manner meet Whenceness, which captures the condition of origination required by free action. Since AC obviously satisfies Whenceness, we have some reason to add to our theory of free action the conceptual primitive of an agent's strictly and literally directly producing some event.

It is worth discussing how considerations from simplicity seem to be what motivate Robert Kane's (1996a) *Free Agency Principle*. Kane (1996a: 122) claims,

> The idea is to see whether such an account can be given for libertarian free agency *without* invoking a special notion of nonoccurrent [i.e., agent-] causation or any other libertarian strategem ruled out by the Free Agency Principle.

Indeed, Kane (1996a: 122, <u>my emphasis</u>) asserts that for nearly every AC theorist, "agent-causation or nonoccurrent causation is specially invoked to make sense of libertarian free agency, and it is just such an appeal that is <u>disallowed</u> by the Free Agency Principle."

At first glance, Kane's latter remark suggests an objection. But what is his Free Agency Principle? And is it self-evidently true? Kane's *Free Agency Principle* states:

> In the attempt to formulate an incompatibilist or libertarian account of free agency…, we shall not appeal to categories or kinds of entities (substances, properties, relations, events, states, etc.) that are *not also needed by nonlibertarian* (compatibilist or determinist) *accounts of free agency…* The only difference allowed between libertarian and nonlibertarian accounts is the difference one might expect—that some of the events or processes involved in libertarian free agency will be indeterminate or undetermined events or processes. But these undetermined events or processes will not otherwise be of categories or ontological kinds that do not also play roles in nonlibertarian accounts of free agency…—the difference being that in nonlibertarian theories, these events or processes need not be undetermined. <u>Such differences as there are between libertarian and nonlibertarian theories should flow from this difference alone</u>, and the task will be to make sense of a libertarian freedom…, given this difference.[34]

It strikes me as infelicitous to compare libertarian theories with compatibilist theories in this manner. Why should one be committed to using only the primitives of a necessarily false account? If compatibilism is false, it is necessarily false. Using the same primitives as a necessarily false account seems no more a virtue of one's theory than using some of the alternative account's false propositions as well.

This problem notwithstanding, the crux of Kane's Free Agency Principle resides in his claim underscored above. But Kane proposes no argument on its behalf. Since there is no plausible principle of rational theory-construction to the effect that one should model one's account of free action after what one believes to be a necessarily false account, it seems that Kane's Free Agency Principle should instead go something like this. "If there is an adequate account of free agency not employing the notion of agent-causation, then do not employ the notion of agent-causation." This would confirm Bishop's insightful point regarding simplicity.

However, the AC proponent may agree with this version of the principle in virtue of a necessarily false antecedent. As we saw in Chapter 5, Kane gives us no good reason to think otherwise. That is, he does not earn his (1996a: 123) claim, "Whatever *can* be done to make sense of a free will satisfying UR can be done without [agent-causation]." I conclude that while considerations from simplicity are powerful, my critic needs to supplement them with an adequate Eventist theory of free action.

---

[34] Kane (1996a: 116, <u>my emphasis</u>).

7.3   Donald Davidson's Challenge

Donald Davidson (1980: 52) argues that AC illuminates nothing about the relationship between an agent and her action. His argument takes the form of posing to the AC theorist a nested dilemma, where every horn implies an absurdity. However, I'll review how O'Connor (1995b: 180-2) sufficiently handles this objection.

To set the stage for Davidson's argument, consider some agent Δ who performs a directly free action A. By definition, A is directly free only if the freedom of A in no way derives from or depends on the freedom of any other action that Δ performs. Intuitively, then, A is basic. For, according to the definition of a basic action, an agent performs a basic action A only if the agent performs no other action A* in order to perform A. Moreover, I see no good reason to think that one's directly free action essentially requires performing some other action *unfreely*. So, for sake of argument, imagine a case where one's directly free action is a basic action.

Davidson's argument is fairly straightforward. We suppose that some agent Δ performs a directly free and basic action A. Davidson contends that the following proposition is obvious:

(DAV)   AC entails that for some agent Δ's directly free (basic) action A, Δ directly produces A.

Seeing how this contention of Davidson's spells trouble for the AC theorist is easy. For, either *Δ's producing A* is an event distinct from A, or it is not. Take the first horn, supposing that *Δ's producing A* is an event distinct from A. Then, either *Δ's producing A* is itself an action or it is not. Surely *Δ's producing A* is an action; otherwise, we have an agent directly causing something without doing anything, which is absurd. However, on pain of contradicting our hypothesis that A is basic, we cannot suppose that *Δ's producing A* is itself an action distinct from A. It follows that *Δ's producing A* cannot be an action distinct from A.[35] Against this consequence, Davidson (1980: 52-3) reasons:

> [T]hen what more have we said when we say the agent caused the action than when we say he was the agent of the action? The concept of *cause* seems to play no role… [N]othing is explained. There seems no good reason, therefore, for using such expressions as 'cause', 'bring about', 'make the case' [or 'produce'] to illuminate the relation between an agent and his act.

---

[35] Incidentally, this is John Thorp's agent-causal position. Thorp (1980: 102) reasons:
> Now presumably we shall want to say that the agent's *causing* the event is also an event. We seem then to have two events, the…alteration in the agent, and the agent's causing the alteration. At once there looms a vicious regress. It can be forestalled only by saying that these apparently two events…are in fact one and the same. This is the logical oddness to which we are committed. … We do not require that an event be the same event as its cause, but that an event be the same as its being caused.

O'Connor (2000a: 58-9 note 33) comments,
> This is simply baffling. Looked at one way—indeed, the only way I am able to get a grip on the claim—it collapses into simple indeterminism: the agent's 'determining' the causally undetermined event is just this latter event's (sic, occurring) (along with there being no prior sufficient [causal] conditions).

The AC theorist may grant the validity of Davidson's argument. However, Davidson provides no reason for his contention that the AC theorist is committed to (DAV). Indeed, Chapter 5 discussed how the AC proponent holds that an agent's directly free action just is her producing a certain event—that is, directly free actions are ACEs. Philosophers besides O'Connor have recognized that an AC theorist may maintain this intuitive claim. For example, Bishop (1989: 68) says:

> [Davidson's] difficulty dissolves, however, once we expose the misunderstanding of the agent-causationist's view that generates it. The theory [of AC] is that actions consist in the causing by their agents *of certain events or states of affairs*. Thus, agents are not held to agent-cause their *actions* (as the constructor of the dilemma implicitly assumes) but rather the events or states of affairs that are, so to say, *intrinsic* to their actions.

Consider also Ginet (1990: 12) who, in explicating AC, observes,

> For the agent-causation analysis to apply, there must be within the simple mental act an event that is only a part and not the whole of it, an event such that its having the extrinsic agent-causal relation to the agent constitutes the whole act.[36]

In addition to the fact that a directly free action may be an ACE, we saw in Chapter 5 that nothing can produce an ACE. Thus, (DAV) is unfounded and believed to be false by some reasonable AC theorists. If (DAV) is false, Davidson's argument cannot get off the ground. His contention is contentious, since an agent cannot agent-cause her directly free action. Indeed, one may view Davidson's Challenge as a reductio against the claim that an agent produces her directly free action. So, rather than being a devastating objection against AC, Davidson's argument further underscores the claim that nothing (including the agent herself) can agent-cause her agent-causing another event.

The next natural step for the Davidsonian critic would be to press the AC theorist on her claim that nothing can produce an ACE. Our critic may complain that if nothing produces an ACE, why not just adhere to *simple indeterminism (SI)*? Why posit agents as strict and literal causes of events? Why not keep things plain and simple, as the simple indeterminist does, helping oneself only to events?

In reply, the objection simply rehearses the objection from simplicity discussed in the last section, §7.2. I concede that if AC and SI are equally satisfactory from a theoretical viewpoint, then one should adopt SI, as it is the simpler view. Nevertheless, I've already provided strong reasons for preferring AC to SI. In particular, I argued in Chapter 5 that while AC meets the condition of origination required by free action, SI has special problems in meeting the condition of origination.

---

[36] Continuing, Ginet (1990: 12-13) opines, "And there is no such event." Now Ginet offers no argument for this latter claim, unless he is assuming that a simple mental act has no internal causal structure, in which case he gives no support for thinking that a directly free action must be simple. Ginet does argue for the less (but still) controversial claim that acting freely does not require an ACE. My point here is that Ginet recognizes that agent-causalists may hold that an ACE itself is the agent's directly free action.

7.4 Multiplying Agent-Causal Events[37]

One might be tempted to criticize AC on the grounds that the view implies the absurd result that an agent is responsible for her agent-causing an event only if there simultaneously occurs an infinite number of ACEs. Of course, the criticism sticks only if our critic clearly shows both that our view should imply a certain result and that this result is absurd. This section undermines confidence one might have in this critique of AC.

Before abandoning agent-causation, Roderick Chisholm maintained that, necessarily, for any agent Δ and any causally simple event e, Δ is responsible for agent-causing e only if Δ is responsible for causing *Δ's agent-causing e*.[38] Consequently,

(1) $\Box\forall\Delta\forall e\{$(Agent Δ is responsible for *($\Delta\rightarrow e$)*) $\supset$ (Δ is responsible for *Δ's causing ($\Delta\rightarrow e$)*)$\}$.[39]

It follows that an agent is responsible for directly producing a causally simple event only if the agent causes, in addition to that causally simple event, an infinite number of causally complex events, where each of these causally complex events occurs precisely at the time the causally simple event e occurs. For example, suppose that Al is responsible for directly producing a causally simple event that is *the coming to be of a state of intention to sip scotch*. Given (1), we infer that Al also simultaneously causes the causally complex event that is *his directly producing the state of intention to sip scotch*. And Al simultaneously causes this second-order causally complex event. For the same line of reasoning, Al simultaneously causes a third-order causally complex event, so on and so forth. Moreover, Al is responsible for each and every one of these higher-order events as well.

Now (1) is an interesting claim and is supposedly at the heart of a devastating objection against AC. If (1) is supposed to figure as an essential premise in such an objection, then we need good reasons for thinking that (1) is true. Moreover, we would need good reasons for thinking that simultaneously causing an infinite number of ACEs is inherently problematic.

It should be clear by now that I already think that producing an ACE is inherently problematic, as I side with O'Connor in holding that nothing might produce an ACE—see

---

[37] I learned a great deal on this topic from Tom Flint's (2005) superb but unfortunately unpublished essay, 'Regresses and the Theory of Agency.'

[38] See, for example, Chisholm (1971: 40-1, 1969: 214). I'll focus on the latter, as it is the most detailed of Chisholm's arguments. However, my points apply just as well to (1969: 214). Chisholm (1976b: 205-7) is also committed to the agent's doing an infinite number of things, but there is no explicit mention of agent-causation. Chisholm (1976a: 69-72) also provides a theory of an agent's causing or bringing about a state of affairs that makes no mention of the agent-causal relation. And Chisholm (1995) flatly denies agent-causation.

[39] Chisholm (1971: 41) suggests that Suarez (1994: Disputation 18) and Robert Binkley (1965) hold this thesis.

Chapter 5, §5.1-5.6. I shall not rehearse this argument again.[40] But if I'm right, then if (1)'s term 'causing' should be interpreted as 'producing,' then since I also believe in the coherency of agent-causation, I have reasons for rejecting (1). On the other hand, if 'causing' should be read as 'merely-causally-contributing', then I again may reasonably reject (1), as I see no good reason on its behalf. Notice that rejecting (1) does not suffice for rejecting my agent-causal view, unless it can be shown that I, *qua* agent-causalist, am committed to (1). The question, then, is whether or not there is a good reason for thinking that my AC view, perhaps when coupled with obvious truths, commits me to (1).

7.4.1   Assessing Chisholm's Case for Proposition (1)

Consider Chisholm's case for thinking that an adequate agent-causal view implies (1). He (1971: 40-1) reasons:

> Suppose, then, that on a certain occasion a man does cause a certain event *e* to happen. What, now, of *that* event—the event which is his thus causing *e* to happen? We have assumed that there is no sufficient causal condition for his causing *e* to happen. Shall we say it was not caused by anything? If we say this, then we cannot [properly] hold *him* responsible for his causing *e* to happen. What we should say, I believe, is that if a man causes a certain event *e* to happen, then, *ipso facto*, he causes it to happen that he causes that event *e* to happen.[41]

We agree with Chisholm that *an agent's directly producing e* is itself an event, *viz.*, an ACE. We agree that there is no sufficient causal condition for an ACE, for we assume incompatibilism. But what else is going on here? How exactly should we understand Chisholm's argument?

Let's turn, then, to see whether we can make Chisholm's argument more rigorous. Once we make his case more rigorous, we'll see where it errs. The more formal argument that follows is my attempt to extract a careful argument from Chisholm's questions and assertions (recall that '*a→b*' designates '*a* produces *b*'). So, as I see it, Chisholm deduces (1) from the following four propositions:

(2)   □ ∀y∀z{(Agent z is responsible for *(z→y)*) ⊃ ~∃x(x is a sufficient causal condition of *(z→y)*)}.

(3)   □ ∀y{∃z(z is responsible for event y) ⊃ ∃x(x causes y)}.

(4)   □ ∀x∀y∀z{((z is responsible for *(z→y)*) & (x causes *(z→y)*) & (x is not a sufficient causal condition of *(z→y)*) ⊃ (x = z)}.

---

[40] One may also think that Davidson's Challenge strikes against the claim that a person agent-causes her directly free action, for her directly free action would not then count as a basic or primitive action. I have no conviction on whether this objection sticks.

[41] This argument also occurs in Chisholm (1976a: 71).

(5)   □ ∀y∀z{((z is responsible for y) & (z causes y)) ⊃ (z is responsible for *z's causing y*)}.

        Together, (2)-(5) strictly imply (1). I affirm the validity of this deduction. Thus, since I deny (1), I'm obliged to deny at least one of these four basic propositions. Before I discuss them, however, let's first work through the deduction to see how (2)-(5) jointly imply (1).

(6)   Suppose that agent Δ is responsible for her ACE, *Δ's producing event e*,
        i.e., suppose that Δ is responsible for *(Δ→e)*. [Assume for conditional proof]

(7)   Thus, ~∃x(x is a sufficient causal condition of *(Δ→e)*). [2,6]

(8)   Thus, ∃x(x causes *(Δ•→e)*). [3,6]

(9)   Thus, suppose that c causes *(Δ•→e)*. [∃-elimination, 8]

(10)   Thus, c is not a sufficient causal condition of *(Δ•→e)*. [7,9]

(11)   Thus, Δ is responsible for *(Δ→e)*, c causes *(Δ→e)*, and c is not a sufficient causal condition of *(Δ•→e)*. [&-introduction, 6,9,10]

(12)   Thus, c = Δ. [4,9]

(13)   Thus, Δ causes *(Δ→e)*. [Substitution, 9,12]

(14)   Thus, Δ is responsible for *(Δ→e)*, and Δ causes *(Δ→e)*. [&-introduction, 6,13]

(15)   Thus, Δ is responsible for *Δ's causing (Δ→e)*. [5,13]

(16)   Thus, Δ is responsible for *(Δ→e)* only if Δ is responsible for *Δ's causing (Δ→e)*.
        [Completes conditional proof, 6-15]

(17)   Thus, ∀Δ∀e{(Agent Δ is responsible for *(Δ→e)*) ⊃ (Δ is responsible for *Δ's causing (Δ→e)*)}. [∀-introduction, 16]

(1)   Thus, □∀Δ∀e{(Agent Δ is responsible for *(Δ→e)*) ⊃ (Δ is responsible for *Δ's causing (Δ→e)*)}. [S5, Rule of Necessitation N: 6-17[42]]

        Again, there are four underived premises—namely, (2)-(5). I affirm both (2) and (5). My reasons for (2) should be familiar by now. I've already argued in Chapter 5 that nothing can produce an ACE. And since a sufficient causal condition of event E clearly produces E, there can be no sufficient causal condition of any ACE. Now, this is not Chisholm's reason for (2). Rather, Chisholm's incompatibilism commits him to (2), which is fine and good.

---

[42] Roughly, the Rule of Necessitation (N) states that every theorem is a necessary truth.

So, I maintain that either (3) or (4) is false, depending on how one interprets the causal relation denoted by the term 'causes'. For, either 'causes' refers to the strong sense of causation, i.e., 'causes' denotes the causal relation of full-fledged production, or else 'causes' refers to the broader and weaker sense of causation, i.e., 'causes' denotes the relation of causal contribution. On the strong reading, (3) is false and (4) is true. On the weak reading, (4) is false while (3) may be true. Either way, there is a false premise and so Chisholm's argument misfires. Or so I contend.

To unpack my argument in more detail, consider the dilemma: either the term 'causes' in (3) and (4) denotes the relation of production or else it denotes the relation of (mere) causal contribution.

Take the first horn. The corresponding readings of (3) and (4), then, are

(3*)   □ ∀y{∃z(z is responsible for event y) ⊃ ∃x(x produces y)}, and

(4*)   □ ∀x∀y∀z{((z is responsible for *(z→y)*) & (x produces *(z→y)*) & (x is not a sufficient causal condition of *(z→y)*) ⊃ (x = z)}.

Although I admit (3*) is intuitive, (3*) is not obviously true. And with careful reflection, we can see that the AC theorist need not accept it. For, as we saw in Chapter 5, an AC proponent may reflect on the nature of special events, *viz*., ACEs, and see that nothing can produce them.[43] Hence, the AC proponent may consider a case where an agent is responsible for her own ACE even though nothing produces this ACE. Furthermore, this is why an AC proponent accepts (4*). More precisely, if one already thinks that nothing can produce an ACE, then one will hold that (4*) is vacuously true due to the impossibility of its antecedent.

The upshot is this: if one is already familiar with my account of AC, one will not be inclined to accept (3*). Thus, if 'causes' denotes the causal relation of production, then Chisholm's argument for (1) fails.

Before moving on to the second horn, permit me to elaborate on my admission that (3*) is intuitive. First, while I find (3*) intuitive, I don't think it is more intuitive than every underived premise I used in arguing that nothing can produce an ACE. Why, though, might one find (3*) so intuitive?

One reason may be that one is not cognizant of the possibility that the universal quantifier ranges over both events with an internal causal structure as well as causally simple events. We—and non-philosophers especially—seldom pay special attention to causally complex events. So, even though most people probably wouldn't put it this way, perhaps people assent to (3*) due to their unreflective belief in the proposition

(18)   □ ∀y{∃z(z is responsible for causally simple event y) ⊃ ∃x(x produces y)}.

---

[43] Again, in order to undermine my argument in Chapter 5, my critic needs to provide a counterexample to the claim that something can *produce* a causally complex event without even contributing causally to the first component or leading edge of that causally complex event. Moreover, there is no substance to any alleged counterexample that invites one to consider something's producing an ACE, for that is precisely what is at issue. The alleged counterexample, then, should rely only on causation by events.

The level of credence for (18) is obviously higher than it is for (3*). For (3*) implies (18), yet (18) does not imply (3*). The intuitive nature of (18)—supposing that is what one normally has in mind when contemplating the alethic status of (3*)—may very well carry over to regarding (3*) as intuitive. I find (18) fairly intuitive. However, it is of no use to my critic, for an ACE is not a causally simple event.

The same sort of point can be made with the following proposition:

(19)   □ ∀y{∃z(z is responsible for event y and ◊∃x(x produces y)) ⊃ ∃x(x produces y)}.

The level of credence for (19) is obviously higher than it is for (3*). Clearly most events we contemplate bear the property of being possibly produced. (3*) implies (19). But the favor is not returned, for there may be events for which someone is responsible that nonetheless cannot be produced by something.

I find (19) very intuitive. But, like (18), it is of no use to my critic. He would need a premise stating that an ACE can be produced. However, we have seen no independent reason for this premise, and we have seen good reasons to reject it.

The crux of this little detour concerning the intuitive nature of (3*) is this. There are other propositions in the neighborhood that are even more intuitive than (3*), and one's confidence in (3*) may unwarrantedly come from the confidence one may have in any one of these other, more intuitive, propositions. Finally, Chisholm's argument for (1) cannot get off the ground if one is restricted to using only these other, more intuitive, propositions.

Now consider the second horn. That is, let's assume that 'causes' denotes the relation of causal contribution. The corresponding readings of (3) and (4) would be

(3**)   □ ∀y{∃z(z is responsible for event y) ⊃ ∃x(x causally contributes to y)}, and

(4**)   □ ∀x∀y∀z{((z is responsible for *(z→y)*) & (x causally contributes to *(z→y)*) & (x is not a sufficient causal condition of *(z→y)*) ⊃ (x = z)}.

(4**) is not obviously true. Indeed, (4**) is obviously false. (4**) implies that every causal contributor of an agent's ACE is the agent herself. That is, only the agent can contribute causally to her own ACE. But this is absurd. Surely there are necessary causal conditions of an agent's ACE, conditions that enable the agent to perform her directly free action. There are a host of environmental conditions that count as events causally influencing a finite agent's directly free action. Typically, an agent's ACE does not happen out of the blue, having no etiology whatsoever.

A brief moment of reflection, then, reveals that an event might merely causally contribute to an agent's directly free activity. Thus, an agent-causalist who believes that an agent's directly free activity is a certain ACE may think that an event merely causally contributes to an ACE. Hence, it is false that *only* an agent can contribute causally to her own ACE. Thus, (4**) is false. As for (3**), perhaps it is true, and perhaps it is false. I'm inclined to think that it is false, strictly speaking, since there may be no causal contributor to God's first free action. Hence, if 'causes' denotes the relation of causal contribution, then again we see that Chisholm's argument for (1) fails, which completes the dilemma. So, either way you interpret 'causes', Chisholm's argument founders.

164

7.4.2 Chisholm's *Tu Quoque* Argument

Before we turn to assessing Donagan's complaint against multiplying ACEs *ad infinitum*, I consider and reject one way Chisholm aims to wriggle free from the putative problem. Chisholm aims to show that multiplying causally complex events is not a special problem for the AC theorist, but rather a problem for deterministic event-causation as well. If Chisholm is right that everyone (except those who deny both agent-causation and deterministic event-causation) is in the same sinking ship, then it is not the case that AC should be rejected on the basis of unnecessarily multiplying events.

Let's turn, then, to Chisholm's tu quoque argument. In replying to one who questions the coherency of an agent's simultaneously producing an infinite number of ACEs, Chisholm (1978: 625-6) says:

> Let us note that such questions may be asked of causation generally—whether "agent-causation" or "event-causation." If the striking of the match made it happen that the building burned down, didn't it also make it happen that the striking of the match made it happen that the building burned down, and therefore also that the striking of the match made it happen that the striking of the match made it happen that the building burned down—and so on *ad indefinitum* (sic)?

To put Chisholm's point least controversially, he thinks that anyone who believes in deterministic event-causation must hold to the following proposition:

(20) $\Box\forall x\forall y\{$(event x deterministically causes y) $\supset$ (x deterministically causes *x's deterministically causing y*)$\}$.[44]

Recall that an event C deterministically causes E precisely when C is a sufficient causal condition of E. Say that 'C scc$\rightarrow$ E' designates 'C is a sufficient causal condition of E.' So, (20) is equivalent to

(20*) $\Box\forall x\forall y\{$(x scc$\rightarrow$ y) $\supset$ (x scc$\rightarrow$ (x scc$\rightarrow$ y))$\}$.[45]

---

[44] I have a hard time believing (20). It seems to fly in the face of *the Principle of Causal Posterity:*

> (PCP) $\Box\forall x\forall y\forall z\{$((x directly causes y) & (z is an essential event-constituent of y)) $\supset$ z is causally posterior to x$\}$.

I take it as obvious that for any causally complex event *x's causing y*, x is an essential event-constituent of the complex event. So, by PCP, x is causally posterior to anything that causes *x's causing y*. But since x cannot be causally posterior to itself, x cannot cause *x's causing y*.

[45] (20*) has an interesting consequence, *viz.*,

> (26) $\Box\forall x\forall y\{$(x scc$\rightarrow$ (x scc$\rightarrow$ y)) $\supset$ ~$\exists z$((x includes z) & (z is a necessary causal condition of (x scc$\rightarrow$ y))$\}$.

Presumably, an event like (C scc$\rightarrow$ E) has many necessary causal conditions. It seems quite odd to think that a particular type of sufficient causal condition of (C scc$\rightarrow$ E) cannot include even one of these necessary causal conditions. Indeed, one might be tempted to think that a sufficient causal condition of

Therefore, Chisholm at least maintains that anyone who believes in deterministic event-causation is committed to (20*). So, there are two questions before us. First, is (20*) true? Second, is the truth of (20*) bad news for a particular critic of AC—the critic who attacks an AC theorist's commitment to (1)? Let's turn to the latter question, seeing what work (20*) does in Chisholm's tu quoque argument.

(20*) implies that for any instance of deterministic event-causation, there will be an infinite number of simultaneous instances of deterministic event-causation. For, suppose the event *(C scc→ E)* occurs. Given (20*), it follows that (C scc→ (C scc→ E)). Again, given (20*), it follows that (C scc→ (C scc→ (C scc→ E))). This procedure may be repeated indefinitely. Hence, there will be an infinite number of simultaneous instances of deterministic event-causation.

Chisholm's contention, then, is that there should be no problem in believing that for some instances of causal production, there are an infinite number of simultaneous instances of causally complex events. It follows that there is a problem in thinking that there is an infinite number of simultaneous causally complex events *only if* the AC theorist who holds to (1) is in no worse shape than anyone else who believes in some event's being deterministically produced. In other words, an AC proponent holding to (1) need not be ashamed in thinking that there is an infinite number of ACEs for each free action, since anyone believing in deterministic event-causation is committed to thinking that there is an infinite number of simultaneous causally complex events. To put the point in yet another way, the possibility of deterministic event-causation and the possibility of an ACE, together with (20*), entail that an infinite number of simultaneous ACEs is impossible *iff* an infinite number of simultaneous deterministic causal transactions is impossible. Thus, since everyone believes there could be deterministic event-causation and since everyone presumably should believe in (20*), then there's no problem in holding that there could be an infinite number of simultaneous ACEs. In fine, (20*) implies that excessively multiplying ACEs is not a special problem for the AC theorist.

The crux, then, concerns the alethic status of (20*). Is (20*) true? At first glance, it seems so, for (20*) appears to be a consequence of the standard account of a sufficient causal condition. How so?

To see how, recall that event C is a *sufficient causal condition* (SCC) of event E *iff* it is not logically necessary that if C occurs, then E occurs; and it is physically necessary that if C occurs, then E occurs either at that same time or later. Now consider a case of deterministic event-causation, supposing that (C scc→ E). By the definition of SCC, C does not logically imply E. However, it is a law of nature that if C occurs, then E occurs either at that same time or later—read as 'L(C ⊃ E).' Now consider whether (C scc→ (C scc→ E)). Obviously, C does not

---

an event E includes at least one necessary causal condition of E. This temptation, more precisely, is to believe

(27)   □∀x∀y{(x scc→ y) ⊃ ∃z((x includes z) & (z is a necessary causal condition of y))}.

Now (26) and (27) are jointly inconsistent. Why? Recall the definition of a necessary causal condition: Event C is a *necessary causal condition* of event E *iff* it is not logically necessary that if E occurs then C occurs; and it is physically necessary that if E occurs, then C occurs either at that same time or earlier. Since (C scc→ E) logically implies C, then (C scc→ E) logically implies anything C logically implies. Hence, nothing C implies can be a necessary causal condition of (C scc→ E). So, (27) implies that (20*) is false. I'm very much inclined to think that (27) is true. And if (27) is true, then Chisholm's tu quoque argument fails.

logically imply (C scc→ E). For, if C did logically imply that (C scc→ E), then since (C scc→ E) logically implies E, C would logically imply E. But we've already assumed that C does not logically imply E. Hence, C does not logically imply (C scc→ E). So far so good. We still need to show that (C scc→ (C scc→ E)). How do we do this?

　　　We can show that (C scc→ (C scc→ E)), *if* we can show that it is a law of nature that if C occurs, then the event *C's being a sufficient causal condition of E* occurs at or later than the time of C. That is, if we can show that L(C ⊃ L(C ⊃ E)), then we can infer that (C scc→ (C scc→ E)).

　　　Do our assumptions thus far imply that L(C ⊃ L(C ⊃ E))? Notice the distinction between L(C ⊃ L(C ⊃ E)) and L(C ⊃ (C ⊃ E)). Our focus is on the former, not the latter. Can we earn the former? Well, we already know that L(C ⊃ E), since (C scc→ E). Is this enough? Since C and E are arbitrary events, what we are really asking is whether the following proposition is true:

(21)　　□∀x∀y{L(x ⊃ y) ⊃ L(x ⊃ L(x ⊃ y))}.

(21) is not obviously true. But (21) does follow from two very plausible assumptions. First, every law of nature is itself a law of nature. Second, whatever is implied by what is a law of nature is also a law of nature. More formally, these two assumptions are

(22)　　□∀x(Lx ⊃ L(Lx)), and

(23)　　□∀x∀y{(Lx & x⇒y) ⊃ Ly)}.[46]

So, again, we know that L(C ⊃ E). By (22), it follows that

(24)　　L(L(C ⊃ E)).

We also know that L(C ⊃ E) logically implies (C ⊃ L(C ⊃ E)), which when coupled with (23) implies

(25)　　L(C ⊃ L(C ⊃ E)).

Now (25) is our desired conclusion. So, (22) and (23) guarantee that L(C ⊃ E) only if L(C ⊃ L(C ⊃ E)).

---

[46] Notice that (23) is distinct from
　　　(23*)　　□∀x∀y{(Lx⇒y) ⊃ Ly}.
(21) follows directly from (23*). (23*) is equivalent to the conjunction of (22) and (23). Also, notice that (23) and (23*) each imply that every necessary truth is a law of nature. If one is inclined to think that this is wrong, one could modify (23) and (23*) accordingly:
　　　(23')　　□∀x∀y{(Lx & x⇒y & y is contingent) ⊃ Ly)}, and
　　　(23*')　　□∀x∀y{(Lx⇒y & y is contingent) ⊃ Ly}.
I'll keep to (23) and (23*). Nothing hinges on this assumption, as everything I say hereafter applies to (23') and (23*') as well.

Recall that the last step of our argument for (20*) was to show that L(C ⊃ E) implies L(C ⊃ L(C ⊃ E)). So, if (22) and (23) are true, we will have shown that (C scc→ E) only if (C scc→ (C scc→ E)). That is, if (22) and (23) are true, then we will have shown that (20*) is true.

Are (22) and (23) true? I don't know. Chisholm is committed to thinking that they are, but to my knowledge he offers no argument for them. While mentioning (22), Chisholm (1981b: 61), neither advocates nor denies it. I cannot bring myself to doubt (23). I'm inclined to deny (22), though I have no compelling argument against it. At the very least, I conclude that the degree to which one is inclined to deny (22) is at least the degree to which one should find Chisholm's (20*) spurious.[47]

7.4.3 Alan Donagan's Criticism

Alan Donagan (1979: 224-5) challenges the coherency of an agent's simultaneously producing an infinite number of ACEs.[48] In contrast, even philosophers who are not friends of AC see no such incoherency. For example, Peter van Inwagen (1983: 135) reports, "[S]o many philosophers are convinced of [agent-causation's] incoherence—on what grounds, I am not clear." Elsewhere, van Inwagen (2000: 13) notes:

> Nor shall I raise questions about the cause of the event "its coming to pass that Reid is the agent-cause of the antecedent brain-event." Again, I think Chisholm has seen what the friends of agent causation should say about the cause of this event, to wit, that Reid was its agent-cause—and was, moreover, the agent-cause of the event "its coming to pass that Reid is the agent-cause of the event 'its coming to pass that Reid is the agent-cause of the antecedent brain-event'," and so *ad infinitum*.[49]

---

[47] O'Connor contests proposition (20) and so is committed to contesting (20*). He (2000a: 58) reasons:

> It is true that *A's causing B* would not have occurred if A hadn't, but this is merely a consequence of the fact that A is a constituent of the more complex event. We are not getting anywhere in understanding what caused a complex event by pointing to its earliest component.

While this objection cannot obviously be translated into an attack on our underived premises for (20), *viz.*, (22) and (23), there is something rhetorically powerful to O'Connor's complaint. Upon considering some instance of deterministic causation (C→E), suppose you ask me what produced it. Suppose I tell you, "Oh, that's easy. C always deterministically causes *C's deterministically causing E*." It seems that I've offered very little by way of an explanation for (C→E). But deterministic causes are supposed to be the sort of things that explain their effects. On the other hand, if Chisholm is right, then maybe the right way to *get somewhere*—as O'Connor suggests is our aim—would be for us to ask, "What *else* produces (C→E)?" So, maybe we can get somewhere in explaining why a causally complex event occurs by pointing to its earliest component. After all, speaking the truth is getting somewhere, and I may be speaking the truth in citing C as the deterministic cause of (C→E). But not all truths are informative. So, despite my perhaps speaking a truth, we're just not getting far enough to latch onto an informative truth, and so we should go farther still by looking for *other* causes of (C→E).

[48] O'Connor (2000a: 58) approves of Donagan's objection.

[49] See also Martha Klein (1990: 95-102).

While Donagan's challenge, if correct, confirms my contention that nothing can produce an ACE (pacé van Inwagen), Donagan's argument does not appear convincing. Let's turn now to Donagan's argument.

Regarding an infinite series of the kind generated by (1), Donagan (1979: 225) reasons:

> And this series will not be a logically unobjectionable infinite series of causes…
> It will be an infinite series of causes *to be* causes: an infinite series of conditions
> of being a cause that must be satisfied before [something causes an event]. Such
> an infinite series of causings to be causes, like all infinite series of grounds to be
> grounds, is logically vicious. A cause is a cause, as a ground is a ground, only if
> all the conditions for its being so are satisfied. But the conditions for anything's
> being a cause can never be satisfied if, by an uncompletable infinite process,
> every such condition must satisfy some further condition.

Donagan's calling the infinite series an infinite process is misleading, as it suggest that the process must take an infinite amount of time to occur. But an ACE lasts only as long as the event directly produced by the agent. Thus, when $(\Delta \bullet \rightarrow e)$ occurs, if $(\Delta \bullet \rightarrow (\Delta \bullet \rightarrow e))$ occurs, then the amount of time it takes for $(\Delta \bullet \rightarrow (\Delta \bullet \rightarrow e))$ to occur is exactly the amount of time it takes $(\Delta \bullet \rightarrow e)$ to occur. This is a perfectly general point; hence, the infinite number of ACEs (if there is an infinite number of them) lasts exactly as long as the causally simple event *e* produced by the agent Δ.

Moreover, and related to this last point, Donagan gives no reason for thinking that the infinite series of conditions must be satisfied *before* the first-order causally complex event occurs. Again, his use of the term 'before' suggests a temporal ordering, but a temporal ordering does not apply to the ACEs generated by (1).

Notice also that Donagan provides no argument for thinking that the infinite series in question is uncompletable. Seeing these points, Chisholm (1979: 372) provides his own response to Donagan's criticism:

> But must [the agent] make these other causal contributions *before* it contributes
> causally to e? Can't we say—to use Donagan's locution—that [the agent] causes
> the one *by* causing the other? It is not as though [the agent] and e were dominoes
> with all these other events falling between them, each such that it must be
> precedent by another. If we use the domino figure, then we should think of the
> other events, not as standing between [the agent] and e, but as attached to the right
> and left of e, so that, when [the agent] falls, it brings down the whole group at
> once.

I think Chisholm's reply, here, casts enough doubt on Donagan's point. While I agree with Donagan that there cannot be an infinite number of simultaneous ACEs for each ACE, I conclude that Donagan hasn't latched onto an outstanding argument for this contention.

7.5   The Timing Problem: Broad, Brody, Honderich, and Ginet[50]

Again, let '*e*' designate the event constituent of an ACE. Several philosophers think that AC inherently has problems in accounting for why *e* occurs at a particular time, given that what produces *e* does not occur at a time.[51] Call the task of explaining the timing of an event *the Timing Problem*. Notice that the Timing Problem is a problem for every account of free action. The question is whether it is a special problem for our AC view.[52]

When introducing how the Timing Problem is supposed to strike uniquely against AC, most philosophers cite C.D. Broad and Baruch Brody. Broad (1952: 131) inquires:

> How could an event [*e*] possibly be determined [i.e., determined by an agent via directly producing the event a lá AC] to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event?

Similarly, Baruch Brody (1969: xix) presses:

> After all, the agent presumably existed a long time before that particular [*e*], so it is not the mere existence of the agent that produces [*e*]. What then causes [*e*] to take place when it does?

Ted Honderich (1993: 41) follows suit, asking:

> But if the originator in the Juliet story was the same from start to finish, why did it cause her decision M4 when it did, rather than at the earlier time of M3 or the later time of the action A? Why wasn't it *always* causing it throughout its entire career?

One might be tempted to think Broad, Brody, and Honderich uncover a deep problem inherent to AC. But how exactly is the argument supposed to go? What are its underived premises? Brody provides no informative premises. Honderich is no more helpful when it comes to giving a claim whose truth-value one can assess. As we see above, Broad proposes a few questions too. However, Broad is a bit more forthright, when he (1962: 131) later asserts:

> Now it is surely quite evident that, if the beginning of a certain process at a certain time is determined at all [e.g., determined by an agent via agent-causing

---

[50] For ease of exposition, I begin anew in numbering critical propositions. For example, "(1)" designates a proposition in §7.4, yet it designates a different proposition here in §7.5.

[51] In addition to those discussed below, see John Bishop (1989: 69), Bernard Berofsky (1987), Randolph Clarke (1993: 193), Kane (1996a: 121), Irving Thalberg (1983: 153-84), and Gary Watson (1982: 10). For one familiar with their versions of the Timing Problem, it should become obvious how my response would handle their versions of the Timing Problem. For O'Connor's response to the Timing Problem, see his (2000a: 74ff) and (1995b: 184ff). Space prohibits outlining his reply and comparing it to mine. Suffice it to say, I discern nothing obviously inconsistent between them.

[52] For O'Connor's reply, see his (1995b: 183-4) and (2000a: 74-6).

the event *e*, which begins the process], its total cause *must* contain as an essential factor another event or process which *enters into* the moment from which the determined event or process *issues*.

So let's begin with Broad, as he provides a substantive claim. It seems that Broad advances the following proposition:

(1)   Necessarily, event *e* is produced at a particular time (or temporal interval) only if an essential factor in *e*'s total cause is another event occurring at a particular time.[53]

Contrary to Broad's rhetoric, his assertion is neither surely quite evident, quite evident, nor self-evident. Of course, there are many people who, like Broad, find (1) pre-reflectively intuitive. However, some people find it pre-reflectively intuitive that God created the cosmos *ex nihilo*—that God directly produced an event being such that no event occurred before it. Perhaps God determined such a first event not by virtue of events that essentially involved God (such as God's having certain mental states). Perhaps God directly caused it by agent-causing it. It would be sheer hubris to think that these people are self-evidently unreasonable in their belief.

Broad, then, has not given us a necessary truth that every reasonable person should accept upon reflection. But are things different for finite creatures such as us, creatures who act in a world already well underway? Maybe so.

The AC theorist, though, may readily admit that for every effected event *e* intrinsic to a finite creature's ACE, *e* has an event that, given the laws of nature, is an essential factor in *e*'s total cause. After all, we know that there are many factors causally influencing and therefore contributing causally to our directly free actions. An ACE may have many necessary causal conditions and so have events that causally contribute to its occurrence. Moreover, these causal contributors, being events, are datable entities.

The AC theorist may consistently maintain that whatever contributes causally to an ACE also contributes causally to the event-constituent *e* of that ACE. She relies on the following self-evident proposition:

(2)   □∀x∀y∀z{(z contributes causally to (x•→y)) ⊃ (z contributes causally to y)}.

So, something may contribute causally to event *e* by merely contributing causally to (Δ•→e). Thus, since events are essentially datable entities, there is no objection against AC to the effect that, necessarily, *e*'s <u>entire</u> etiology, though *e* is produced by an enduring substance, excludes all events.

We now can see that the AC theorist has sufficient resources to answer the question, "How can an agent Δ be capable of producing *e* now, when Δ was incapable of producing *e* a few moments ago?" Easy. Perhaps all of the necessary causal conditions for (Δ•→e) did not occur a few moments ago, but they occur now.

We may likewise answer Honderich, who invites us to explain why, for example, Juliet *qua* agent-cause causes some event *e* at moment M4 rather than at an earlier moment M3 or at a

---

[53] More formally, □∀x{(event *e* is produced at a particular time or temporal interval) ⊃ ∃y∃z((y is x's total cause) & (z is an essential factor in y) & (y is another event occurring at a particular time))}.

later moment M5? "Why wasn't [Juliet] *always* causing [*e*] throughout [her] career?" First, according to our view, an agent directly produces *e* precisely when *e* occurs. Second, every necessary causal condition for the ACE *(Juliet•→e)* occurs at M4, but perhaps some necessary causal condition of *(Juliet•→e)* neither occurs at M3 nor occurs at M5. In short, the Timing Problem is not a special problem for the AC proponent.

Brody (1969: xix) inquires, "What then causes *e* to occur when it does?" Well, the what-question is easy. The short answer is *the agent*. And this is not to say that the mere existence of the agent causes *e*. And since the agent causes *e* to occur, and since *e* can only occur when it does occur, it follows that the agent is *that which* causes *e* to occur precisely when *e* occurs. *Why* does *e* occur precisely when then as opposed to some other time? Because the agent Δ directly produces *e* precisely at the moment *e* occurs. The ACE is not the same event as its event-constituent, *e*. And an ACE, being an event, is as much a datable entity as any other event. William Rowe (1991b: 248) notes,

> It is incorrect, therefore, to suppose that when a person agent-causes [*e*] the only event to be found in the neighborhood of the agent is [*e*]. Without the exercise of active power [i.e., the ACE] there can be no occurrence of [*e*].

Now one might think that the AC theorist is simply pushing the problem back a step. For the critic may ask about the timing of an ACE. Since nothing can produce an ACE, *a fortiori* there is nothing that produces the ACE that explains the timing of the ACE. Nevertheless, an ACE may have mere-causal-contributors—e.g., a necessary causal condition. And since the timing of a mere contributor provides some explanation of the timing of its effect, the AC theorist may help himself to this fact as much as anyone else in explaining the timing of an agent's directly free action. For example, an AC theorist can easily answer why a particular ACE occurs now rather than one year ago, even though the agent existed a year ago. For, there may be many necessary causal conditions of the ACE's occurring now that did not occur until quite recently. So, since there are events in the immediate etiology of an agent's ACE, there is room for thinking that their timing sufficiently explains the ACE's timing. And, explaining an ACE's timing suffices for explaining the ACE's event-constituent.

What would a more formal version of such an explanation look like? Say that x *forthwithly* contributes causally to y *iff* x contributes causally to y and x does not contribute causally to any z that contributes causally to y.[54] Our critic presumably assumes that the timing of event *b* may be explained by the timing of event *a* in virtue of *b*'s occurring immediately after *a*, i.e., *a* antecedently directly produces *b*. The AC theorist analogously maintains that every ACE, e.g., (Δ•→e), occurs immediately after every event that antecedently forthwithly causally contributes to (Δ•→e). If our critic assumes that simultaneous production of an event is possible, then our critic may also
assume that the timing of *a* explains the timing of *b* in virtue of *b*'s occurring simultaneously with *a*, i.e., *a* simultaneously directly produces *b*. In this case, their timing would be the same. The AC theorist may analogously maintain that simultaneous causal contribution is possible, claiming that every ACE *(Δ•→e)* occurs simultaneously with every event that simultaneously forthwithly causally contributes to (Δ•→e).

---

[54] I would use the relation *x's directly causally contributing to y*, had I not used it to define another relation in Chapter 6.

Finally, let's consider Ginet's (1990: 13) confident formulation of the Timing Problem:

> More decisive is the difficulty…that if the cause of the mental occurrence is just me, just the enduring entity, and no event at all, then it cannot explain what it needs to explain. A merely enduring thing as cause lacks the features needed to make it capable of explaining the particulars of the mental occurrence. It cannot, for instance, explain its timing.[55]

Notice that Ginet, in keeping with his *Simple Indeterminism*, does *not* assume that the timing of every event *must* be explained. Rather, he stipulates that a cause must explain the timing of its product. More precisely, Ginet assumes the proposition that

(3) $\Box\forall x\forall y\{(x\bullet\rightarrow y) \supset (x$ explains the timing of $y)\}$.

Ginet provides no argument for this claim. Moreover, the AC theorist may explain the timing of an event *b* by referring to the timing of a mere-causal-contributor *c* of $(a\bullet\rightarrow b)$. More precisely, the AC theorist may hold to the following proposition:

(4) $\Diamond\exists x\exists y\exists z\{(x\bullet\rightarrow y)$ & $(z$ forthwithly contributes causally to $(x\bullet\rightarrow y))$ & $(z$ explains the timing of $y)$ & $\sim(x$ explains the timing of $y)\}$.

Someone who already holds to (4) has no motivation to believe (3). Indeed, (4) implies that (3) is false. Moreover, the AC proponent is not committed to there being an event whose timing is essentially inexplicable.

Suppose that the critic of AC retrenches, proposing as an underived premise the proposition that, necessarily, every part of the etiology of *e* is an occurrence at a particular time. Randolph Clarke (1996a: 40-1) correctly responds, "However, this modified objection is simply the assertion, without further reason, that agent causation is impossible."

Interestingly, an AC proponent is committed to believing that every event that forthwithly contributes causally to an ACE *ipso facto* forthwithly contributes causally to the event-constituent of the ACE. That is, AC theorists are committed to the following proposition:

(5) $\Box\forall\Delta\forall x\forall y\{($Agent $\Delta\bullet\rightarrow x)$ & $(y$ forthwithly contributes causally to $(\Delta\bullet\rightarrow x)) \supset (y$ forthwithly contributes causally to $x)\}$.

Thus, an AC advocate believes that *e*'s timing may be explained by the timing of some event *c* that forthwithly contributes causally to *e*. For, along with everyone else, an AC proponent believes the following proposition:

(6) $\Box\forall x\forall y\{($event x forthwithly contributes causally to $y) \supset (y$ occurs at or right after the time of $x)\}$.

---

[55] My response applies equally well to Ginet's (1997: 94) twist on the Timing Problem. For O'Connor's responses, see his (1995b: 183-4) and (2000a: 74-6).

Again, an AC theorist should concede that, for finite creatures like us, there typically are events that forthwithly contribute causally to an agent's ACE. So, why did the ACE occur right then? Well, look at the timing of any of the events forthwithly contributing causally to the ACE. Consider one of these forthwith-causal contributors and call it C. One might ask, "Why did C become a causal contributor of the ACE right then?" Well, look at the timing of any of the events that forthwithly contributes causally to *C's contributing causally to the ACE*, and surely there are some of these. This strategy may be applied indefinitely. So, it seems that with respect to addressing the Timing Problem, AC is in no worse shape than any other theory of free action.

Now, it should be admitted that the AC theorist denies the following proposition:

(7) $\Box\forall x\forall y\{$(x explains the timing of y by occurring at or right before y) $\supset$ (x deterministically causes y to occur at or right after x).

Indeed, everyone who thinks that the timing of an undetermined free action even might be explained should reject (7). For, (7) simply says that the timing of event is explained only if this event has a direct sufficient causal condition. If a solution to the Timing Problem implies (7), then, again, the Timing Problem is not a special problem for the AC theorist. It would also be a problem for the Simple Indeterminist and the Causal Indeterminist. O'Connor (1995b: 193, <u>my emphasis</u>) observes:

> There will often be certain considerations or other factors at the time of acting that elicited my action (by suggesting that this was a particularly opportune time to satisfy the desire), and these will certainly figure in a full explanation of my action. By the same token, it's not obvious that there *needs* to be <u>such</u> environmental stimuli. Perhaps I am only concerned that I act within a certain time frame, and any particular moment is as good as any other. In such a case, there may not be an explanation of why I acted just *then* (rather than at some other time) [as (7) prescribes].[56]

Moreover, (7) seems to strike against the compatibilist as well. For, suppose compatibilism is true. Now consider an agent who freely performs A, where A is deterministically and directly caused by C. Clearly, the timing of C explains the timing of A. But, we may ask, what explains the timing of the event *C's deterministically causing A*? Suppose that the compatibilist points to C2, since C2 deterministically and directly causes C.[57] Obviously, the timing of C2 explains the timing of C. This may go on indefinitely—C2's timing may be explained by its deterministic cause, C3, and so on and so forth. Nevertheless, it need not go on indefinitely. And suppose it does not. Suppose, that is, that there is some indeterminism in the etiology of C2. Then, according to (7), we will not be able to explain the

---

[56] Of course, for finite agents like us, it seems that there needs to be some or other environmental stimuli. O'Connor is suggesting that the stimuli need not be in the form of factors the agent believes suggest a particularly opportune time to act.

[57] C2, even though it deterministically causes C, may or may not suffice for explaining *C's deterministically causing A*. Let's assume, though, that it does suffice. If not, adjust the argument according, replacing the appropriate occurrences of 'C' with '*C's deterministically causing A*.'

timing of C2. And if we cannot explain the timing of C2—more precisely, if C2's timing cannot be explained in the manner prescribed by (7)—, then we should not be so quick to demand from the outset that A's timing should be explained in the manner prescribed by (7).

In fine, the Timing Problem does not seem to be a special problem for the AC theorist. Everyone must deal with the Timing Problem, and the AC theorist may solve the Timing Problem in much the same way as any other libertarian does. With the Eventist, the AC proponent believes that the timing of an event E may be explained by the timing of an event C that forthwithly contributes causally to E. Moreover, if a solution to the Timing Problem requires (7), then, again, there's no special problem for AC. For, (7) undermines every libertarian theory. I say, "So much the worse for (7)."

Although Randolph Clarke explicitly holds that O'Connor's view of AC cannot handle the Timing Problem, we *may* still agree with Clarke's (1993: 194) line of thought that

> the occurrence of certain prior events will be [given the laws of nature] a necessary condition of an agent's causing a certain event. Absent those prior events, the later event will not be naturally possible, and an agent can cause only what is naturally possible. The agent-causal view thus has the same resources as does a wholly event-causal view of human agency to explain why an agent performs a certain action at a certain time, rather than earlier or later. If there is an event, such as her acquiring new reasons, that explains why she acted then and not at some other time, then both sorts of views have available an explanation. If there is no such event, then neither sort of view has available an explanation.

The moral is that there is more to an adequate theory of AC than simply positing a relation of direct causal production that an agent bears to some event. The AC theorist has more to say. For example: (i) I've detailed how an ACE might have causal contributors, (ii) my view implies that an agent's directly free action just is an ACE, (iii) there might be causal influences of one's directly free action, and (iv) the timing of a causal contributor of an ACE explains the timing of the ACE's event-constituent. Hence, if there is a damaging objection that applies uniquely against AC, it is not the Timing Problem.

7.6   Kane's Case Against Agent-Causation

The strongest case Kane has for denying AC consists of proposing an adequate theory of free action that does not imply the occurrence of an ACE. If he's right, then we should reject AC on the basis of our previous considerations from simplicity. However, we saw in Chapter 5 that Kane does not satisfy the condition of origination that free action requires. Hence, he does not provide an adequate Eventist theory of free action. Nevertheless, in addition to his attempt to offer an Eventist theory of free action, Kane criticizes AC on other grounds. This section discusses how his criticisms fall short.

7.6.1   Paraphrase

Kane suggests an objection based on the idea that something's causing an event might be translated without remainder into events. Kane (1995: 118) claims that AC entails the condition that the agent's "causation of her free choices or actions cannot be explained as the causation of

events or occurrence by other events or occurrences." As a point of clarification, notice that Kane suggests that AC entails that the agent directly produces her directly free action. However, a directly free action may just *be* agent's directly producing a certain event. So let's adjust the AC condition to read "An agent's directly producing an element/event intrinsic to her directly free activity cannot be explained as the causation of events by other events." Kane (1995: 118) then objects:

> This…condition is problematic for a well known reason: while causation by things or substances is common, it can usually be interpreted as the causation of events by other events. "The stone broke the window" is elliptical for "The stone's striking the window caused the window to break."

I fail to see how this is a telling objection against AC. Kane presumably offers a reason here for the condition's being problematic, but it is not a good reason (even if well-known). Consider the following two propositions. First, causation by things or substances can usually be interpreted as the causation of events by other events. Second, causation by things or substances can always be interpreted as causation of events by other events. The former claim does not entail the latter claim. A reasonable AC theorist may hold to the former claim while denying the latter claim. Kane advances no further argument for the latter claim.[58]

7.6.2 Plurality Questions
Elsewhere, Kane (1996a: 121) objects that AC does not

> …help to explain what most needs explaining for indeterminist theories of freedom, namely how the plurality conditions are satisfied: Why did the agent rationally and voluntarily do A here and now rather than doing otherwise? Other libertarian theories […] try to answer this question by citing the agent's reasons, motives, volitions, or other mental states or changes. But agent-cause theories deny that adequate answers to plurality questions can be given in terms of *occurrences* alone of any kinds, physical *or* psychological, past *or* present, involving the agent or not.

An AC theorist may think that there simply cannot be an adequate answer to plurality questions for an undetermined free action; hence, that's why she may deny that there is an adequate answer. A great deal depends on how one formulates what satisfying the plurality conditions requires. Kane's 'rather than' locution suggests that the issue is one of contrastive explanation. Some philosophers think that contrastive explanation requires that the explanandum-event is determined. For example, Clarke (1996b: 189) notes:

> It is sometimes said flatly that when an event is not causally determined, it cannot be contrastively explained. Perhaps what lies behind this rejectionist position is

---

[58] See also Martha Klein's (1990: 101) criticism of this objection.

the view that to explain contrastively is to cite an event or condition given which the outcome had to occur.

That is, some philosophers believe

(1) $\Box\forall x\forall y\forall z\{$(x explains why y occurred rather than z) $\supset$ (x deterministically causes y, and z fails to occur).

Now if (1) is true, then there simply cannot be a contrastive explanation for a free undetermined action. Hence, (1) implies that a contrastive explanation is not a special problem for AC; rather, it is a problem for any libertarian.

Clearly, then, (1) can't be what Kane has in mind, for Kane is a libertarian. Kane rejects (1), believing that there is a free action and that every free action is essentially undetermined.[59] But Kane still thinks that the libertarian can satisfy the plurality conditions for an undetermined free action.

So, what exactly are Kane's plurality conditions? The fine details need not concern us. While further reviewing the nuances of Kane's theory would take us too far afield, Kane clearly holds that indeterminist theories of free action satisfy the plurality conditions by citing certain events in the etiology of the free action.[60] Unfortunately for Kane, he does not show that an AC proponent cannot satisfy his plurality conditions by likewise appealing to events in the etiology of an ACE.

If certain events contribute causally to an ACE, then an AC theorist may appeal to them in much the same way that Kane appeals to events that, according to his theory, produce an agent's undetermined free action. Therefore, Kane's last claim in the block-quotation above is not obviously correct. It seems that whatever answer Kane gives to his plurality questions vis-à-vis an Eventist's candidate for a free action can be translated into an agent-causalist's answer to Kane's plurality questions vis-à-vis an ACE.

Let's see how Kane's argument fails from another angle. Consider, again, Kane's first claim that AC does not help to explain what most needs explaining for indeterminist theories of freedom, namely how the plurality conditions are satisfied. Let's assume that we know what it would take to satisfy the plurality conditions. A question remains. Is Kane really right that explaining how his plurality conditions are satisfied is what *most* needs explaining for indeterminist theories of freedom? Suppose that we concede that such an explanation is important and that success on this front would be a fine achievement. Are there not other important conditions that an adequate theory of free action should meet? For example, what about the origination condition? Would not meeting the challenge of Whenceness be just as important for a theory of freedom as satisfying the plurality conditions? Of course it would, because free action requires the condition of origination. An adequate account of free action requires meeting Whenceness, and we've seen that AC is the best theory for meeting Whenceness.

Let's see how this visceral agent-causal reply translates into a more careful response to Kane's objection. We first get clearer about how Kane's objection is supposed to work. For

---

[59] Moreover, (1) may be false for independent reasons—see Clarke (1996b: 188ff).

[60] Kane (1996a: 28, 35, 73, 113). For the details of Kane's theory, see my Chapter 5.

sake of argument, let's assume that satisfying the plurality conditions can *only* be given in terms of occurrences/events alone, which is an implicit but essential premise in Kane's argument. The issue, then, concerns the relationship between free action and the plurality conditions. It seems that the plurality conditions are supposed to be either necessary or sufficient for free action. Let's take these in turn.

Suppose that the plurality conditions are necessary but not sufficient for free action. This confirms Kane's (1996a: 121) claim that they are what "most needs explaining for indeterminist theories of freedom." Kane's argument, it would seem, would proceed as follows. Free action entails the plurality conditions. Only events can explain the plurality conditions. Hence, free action entails the denial of AC.

But this argument is invalid. To make it valid, one would need to claim that only Eventism can account for the plurality conditions. That is, the plurality conditions entail Eventism. But Kane offers no good argument for this claim. And it's easy to see why he couldn't provide such an argument. For, it might be that free action entails the plurality conditions *and more besides*. For instance, free action entails the condition of origination. We have good reason to think that the condition of origination implies the falsity of Eventism. Thus, even if the AC theorist grants that the plurality conditions can only be accounted for by events, Eventism does not follow.

Suppose that the plurality conditions suffice for free action. It would seem that Kane's objection would proceed as follows. Since the plurality conditions can only be accounted for by events, and since the plurality conditions imply free action, then we can account for free action without employing the notion of AC. Hence, given our considerations from simplicity, we should reject the AC view.

The problem, though, is that Kane offers no good reason for thinking that the plurality conditions suffice for free action. It is not obvious that they do. And if they do not, an agent-causalist may concede that adequate answers to the plurality questions might be given in terms of occurrences alone.

In addition, the plurality conditions seem to be insufficient for free action. An unfree action might satisfy the plurality conditions. But an AC theory need not offer a special or unique explanation of unfree action. The AC proponent is in the business of accounting for free action. Moreover, the incompatibilist has a reason to think the plurality conditions do not suffice for free action. For example, even if determinism is true, one can account for why an agent rationally and voluntarily does A here and now rather than doing otherwise. However, the agent's performing A would not be free, as it would be determined.

One should take care to note that the signature feature of the AC view—*viz*., positing an ACE—meets Whenceness. However, a signature feature *of* a view need not exhaust the view. There is more to our AC view than simply positing an ACE. If free action requires satisfying the plurality conditions, then an AC theorist may use other resources at her disposal, and these other resources may or may not be cashed out in terms of events alone. Again, the AC theorist may explain how the plurality conditions *proper* (supposing they do not imply free action) are satisfied in much the same way as any other libertarian would. An AC theorist, for example, may speak of the agent's reasons, motives, volitions, or other mental states or changes as influencing the agent's directly free action. Here, the influence may be taken as primitive (as a Simple Indeterminist may do), or perhaps the influence is a species of causation—perhaps the influence will be mere causal contribution rather than full-fledged production.

If it can be shown that the plurality conditions imply free action, the AC proponent will surely deny that one can satisfy the plurality conditions in terms of events alone. For then the plurality conditions would imply meeting Whenceness, and we've reason to think this can't be done with events alone. Similarly, it seems that the only point to satisfying the plurality conditions is to do so *for a directly free action*. But again, on our view, a directly free action is an ACE, and an ACE does not consist of occurrences alone. *A fortiori*, satisfying the set of plurality conditions *for a directly free action* will not be done in terms of occurrences alone.

7.6.3 Explanatory Impotence
Kane (1996a: 188) reasons:

> [T]he question is whether this postulated additional [agent-causal] relation explains whatever it is that cannot otherwise be explained in terms of occurrences or events involving agents—or whether it explains anything at all.
>   Yet agent-causalists are in a bind when it comes to saying something more positive about what agent-causation is or how it operates. For they cannot say what is distinctive about it or its operation in terms of occurrences or events of any kinds, including states or changes involving the agents, physical *or* psychological.

Kane's suggestion, here, seems to be that employing the notion of agent-causation could add nothing uniquely informative to an adequate account of free action. That is, if it added anything informative, then what it adds could be unpacked in terms of events, in which case AC would not be adding anything *uniquely* informative. However, if AC adds nothing informative, then an adequate account of free action could just as well do without the notion of AC. Therefore, given considerations from simplicity (discussed in §7.2), we should not employ the notion of AC.

Since the signature feature of our AC view is the positing of an ACE, an ACE should be the candidate for what adds something uniquely informative to our theory of free action. Of course, I contend that it does. And we can see that Kane's reason for claiming that agent-causalists are in a bind is mistaken. For an agent-causalist *can* say what is distinctive about AC, *and she can even do it in terms of events*.

What is so distinctive about AC? Answer: a directly free action occurs only if—and here's the event—it comes to pass that an agent strictly and literally, *qua* substance and *qua* enduring entity, directly produces an event. Notice that this is not a mainly negative thesis. It is no more a negative thesis than the Eventist's thesis that there is *no* agent-causation involved in an agent's performing her directly free action. So, what is distinctive about AC is the fact that a very special sort of event—*viz.*, an ACE—occurs. For, an ACE is an event, an ACE is uniquely informative in that it meets the challenge of Whenceness. However, none of this implies Eventism. Indeed, an ACE is inconsistent with Eventism.

Kane (1996a: 188, 189) asks rhetorically:

> So the question is, what are we *adding* to all that we can say about the agent's physical and psychological circumstances by adding that…the agents nonoccurrently cause their undetermined free actions [or willings or intentions]?

… What, then, does nonoccurrent causation [i.e., agent-causation] explain about undetermined free actions that cannot be otherwise explained?

In reply, AC answers the following questions that Eventism does not obviously answer. How is an agent a source but not a wholly derived source of an element essential to her directly free action? Whence comes the entire event-constituent of an agent's directly free action? The answer are these, respectively. The agent directly produces the element. The agent.

Kane (1996a: 195) asserts:

In order to provide illumination about the…production of actions [or events intrinsic to actions] by these [agent-causal] entities, [AC theorists] have to say something about the conditions of the agents and what is going on involving the agents when the actions occur. And this is the sort of information that, by its nature, nonoccurrent causation [i.e., agent-causation] does not supply.

Kane is unclear about what precisely requires illumination. The agent-causalist supplies some information about the production of an event intrinsic to a directly free action. She points to the agent-causal *event*.

But perhaps Kane demands from the AC theorist more exposition on what happens in the moments surrounding an agent's directly free action. And presumably he thinks that simply postulating an ACE does not supply this information. But we may concur with this insight, while noting that there is more to an adequate AC view than *simply* postulating an ACE. For instance, I have given a theory of causal contribution. Citing events that causally contribute to an ACE is saying something about the conditions of the agent and what is going on involving the agent prior to her directly free action. So, the agent-causalist may supply all manner of information about what contributes causally to the ACE in much the same way that an Eventist supplies information about what contributes causally to his candidate for a directly free action. As a bonus, the AC theorist meets Whenceness, which Kane's theory should but unfortunately does not satisfy.

7.6.4   Concluding Kane
In conclusion, early in his book Kane (1996a: 14) states confidently:

The air is cold and thin up there on Incompatibilist Mountain, and if one stays up there for any length of time without getting down the other side, one's mind becomes clouded in mist and is visited by visions of noumenal selves, nonoccurrent causes [i.e., agent-causes], transempirical egos, and other fantasies.

Kane offers no good argument for thinking that agent-causation is a fantasy. Some visions are beneficent and aid in burning away mist. Such is the vision of agent-causation. Or so I maintain. We have yet to see how agent-causation is clearly mistaken.

7.7 Van Inwagen

Van Inwagen denies that AC shows how an undetermined free action is possible.[61] That is, he denies that the AC proponent successfully answers the Intelligibility Question. But does van Inwagen offer good reasons for denying that AC nonvacuously and obviously implies that there is an undetermined free action?

I see two ways to earn van Inwagen's critical denial. First, one might argue that AC is metaphysically impossible. If it is impossible, it will vacuously imply anything, which is bad news for the agent-causalist. Second, one could argue that even if the agent-causalist's view is possible, it does not *obviously* imply that there is an undetermined free action. One could accomplish the latter in two ways. First, one could argue that even if AC is possible, its possibility is mysterious. And if its possibility is mysterious, it will not *obviously* imply that there is an undetermined free action. Second, one could argue that the AC account of an undetermined free action is consistent with there being no undetermined free action. This section discusses why the AC theorist should not find van Inwagen's explicit position threatening.

Does van Inwagen provide any good reasons for thinking that AC is metaphysically impossible? No. Indeed, so far as I can tell, he maintains that there is no good argument for its being impossible. I'll note three texts in support of this claim. First, van Inwagen (1983: 135) discloses that "so many philosophers are convinced of [agent-causation's] incoherence—on what grounds, I am not clear." Second, van Inwagen uses the possibility of agent-causation as a premise in an argument against the claim that the thesis of determinism is equivalent to the Principle of Universal Causation. The Principle of Universal Causation is "the thesis that every event (or fact, change, or [obtaining] state of affairs) has a cause."[62] Van Inwagen (1983: 4) reasons:

> Suppose, that is, that the Principle of Universal Causation and…the doctrine of *immanent* or *agent* causation is true. Suppose, to be more specific, that a certain change occurs in an agent, Tom, and Tom himself is the cause of this change, and no earlier state of affairs necessitated this change. Then the thesis of determinism is false. But our description of this case is internally consistent, for it does not entail that any event is without a cause.

Hence, van Inwagen should stand by the soundness of this argument only if he holds that AC is coherent.

However, perhaps he has since changed his mind, which brings us to our third text. Van Inwagen (2001: 22) reports having an inclination to agree with those philosophers who think that agent-causation is either incoherent or metaphysically impossible. However, he (2001: 22) parenthetically notes that he has no firm opinion on this question. It seems, then, that he is committed to having no opinion on whether his argument (just quoted above) is sound. All in all, van Inwagen advances no reasons for thinking agent-causation is impossible.

Does van Inwagen provide reasons for thinking that agent-causation's possibility is mysterious? He (1993: 194-5) reports:

---

[61] See van Inwagen (2001, 2000, 1993: 193-7).

[62] Van Inwagen (1983: 3).

I do not understand agent-causation. At least I don't think I understand it. To me, the suggestion that an individual thing, as opposed to a *change* in an individual thing, could be the cause of a change is a mystery. … I am saying only that agent-causation is a mystery and that to explain how it can be that someone can have a choice about the outcome of an indeterministic process by an appeal to agent-causation is to explain a mystery by a mystery.

So van Inwagen himself finds the possibility of AC mysterious. However, he offers no reason for thinking that *other* reasonable philosophers should find it mysterious. Therefore, one should be loath to find merit in van Inwgen's claim that appealing to AC to explain X is to explain X by a mystery. For we have yet to be given a cogent argument to think that AC is mysterious. Our objector, if she wishes to take the baton from van Inwagen, needs to propose a necessary feature of being obscure or of being mysterious, show that an adequate theory of free action clearly precludes this feature, and show clearly that our agent-causal view uniquely has this feature.

To be fair, van Inwagen never argues that the possibility of AC is mysterious. However, he explicitly argues that AC does not obviously imply that an undetermined free action is possible. And if his argument succeeds, then the agent-causalist is in trouble. Van Inwagen (2000: 1-2) claims:

I will…present an argument for the conclusion that free will and indeterminism are incompatible even if our acts or their causal antecedents are products of agent causation. I see no way to respond to this argument. I conclude that free will remains a mystery—that is, that free will undeniably exists and that there is a strong and unanswered *prima facie* case for its impossibility.

The argument of which he speaks is a sort of indirect argument against AC. His argues that an undetermined free action is itself impossible. He reasons that since an undetermined free action is impossible, AC of course does not imply the possibility of an undetermined free action. Nevertheless, van Inwagen's argument for the impossibility of an undetermined free action is inadequate, as I carefully criticized this very argument in Chapter 4. Van Inwagen, then, submits no argument against AC that withstands a reasoned evaluation.

7.8   Labels
Some philosophers believe that AC simply *labels the problem* facing libertarians. Others maintain that AC simply labels the libertarian's *solution*. Let's take these in turn.
John Bishop (1989: 69) asserts:

[M]ere appeal to agent-causation is no help in explaining how actions differ from other behavioral episodes that are not 100 percent causally determined. All the notion of agent-causation can do is…*label* what is problematic about fitting actions into the natural causal order by emphasizing our commonsense understanding of actions as controlled exercises of the agent's powers.

I wish to make two points in response to Bishop's claims. First, as we saw in Chapter 5, Bishop's first claim is not obviously correct. On our view, an ACE is a directly free action that cannot be produced. Thus, an ACE is intrinsically such that it cannot be 100 percent causally determined. But this is not obviously the case for Eventist candidates for a directly free action. Hence, an appeal to AC *does* help in explaining how actions differ from other behavioral episodes that are not 100 percent causally determined. Second, Bishop provides no argument for thinking that AC can only label what is problematic for the libertarian. His claim is only a declaration of faith and, absent a cogent reason to the contrary, need not be taken seriously by every reasonable agent-causalist.

In contrast to labeling the *problem* facing libertarians, other philosophers hold that AC simply labels the libertarian's *solution*. For example, Gary Watson (1982: 10) claims, "'Agent-causation' simply labels, not illuminates, what the libertarian needs." Similarly, Kane (1996a: 192) states, "At best, nonoccurrent causation [i.e., agent-causation] seems to be a 'label' for what is wanted…and not an explanation; it does not supply the filling to the pie."

I wish to make two points in response to Watson and Kane's claims. First, if AC labels the libertarian's solution, then there's no sense arguing against it, for one should not disagree with solutions. However, since there is disagreement, there is reason to think that AC entails a proposition contested by other reasonable libertarians and therefore should not be considered a mere label of the libertarian's solution.

For example, the AC theorist makes the following substantive claim. If there is free action, then an agent—*qua* substance—strictly and literally directly produces an element essential to her directly free action. Now, if postulating an ACE is a label for an adequate account of free action (or for a condition required by free action), then everything falling under the label should be implied by any adequate account of free action. Thus, if AC labels an adequate account of free action (or labels a condition required by free action), then Kane's attempt, e.g., to offer an adequate account should also imply that a substance strictly and literally directly produces some event. But Kane's account is Eventist, and Eventism precludes an ACE. Thus, either AC does not label any part of the libertarian's solution or else Kane's account is inadequate. The moral is this. No one who believes in the possibility of libertarianism should both deny the AC view and claim that AC labels an essential feature of an adequate account of free action.

But perhaps I am missing the point of Watson and Kane's objection, and this brings us to my second point. Their remarks also imply that AC neither illuminates nor explains what the libertarian needs or wants. But they offer no good argument for this claim, so we need not take it more seriously than their simply stipulating that AC is explanatorily impotent. They would do better to first propose a condition that free action obviously requires and then argue that our theory of AC obviously precludes it. In the absence of such a reasoned and careful treatment, a reasonable agent-causalist need not find their bold assertions threatening.

7.9   Knowability
Some philosophers criticize AC on epistemic grounds. Consider John Thorp (1980: 101, <u>my emphasis</u>), who reasons that the agent-causalist

> must say that what is means for an agent to cause an event is (a) for that event to
> fail to be explained by causal links to preceding events and (b) for that event to be

linked to the agent by a causal power. It is the presence of (b), the exercise of causal power, that differentiates agent causality from mere uncausality, freedom from mere randomness. <u>But if indeed causal powers and their exercise are not observable ingredients in the world, then we seem to have the highly embarrassing result that we can never know when they are present and when not</u>…

I have two points. The first point is orthogonal to Thorp's central point but still bears on his claims: a reasonable AC theorist may deny Thorp's (a). On our view, there could be an event C that merely causally contributes to an ACE, $(\Delta\bullet\rightarrow e)$, where e occurs later than C. As we noted above when discussing the Timing Problem, the following proposition seems self-evident:

(1) $\Box\forall x\forall y\forall z\{(z$ contributes causally to $(x\bullet\rightarrow y)) \supset (z$ contributes causally to y)$\}$.

Hence, event C could contribute causally to e, even though the agent Δ directly produces e. Since C could precede and contribute causally to e, C could at least partly explain e. Therefore, contrary to Thorp's (a), it need not be the case that e fails to be explained by causal links to preceding events.

My second point concerns Thorp's assertion, underscored above. Incidentally, Thorp is an AC theorist. However, he (1980: 101ff) reasons that an agent's causal powers and their exercise are observable ingredients of the world.[63] But why should one believe Thorp's assertion anyway? That is, why should we think that if causal powers and their exercise are not observable ingredients in the world, then we can never know when they are present and when not? If this is true, it would seem to be true only in virtue of the following, more general, claim:

(2) $\Box\forall x\{(x$ is not an observable ingredient of the world) $\supset$ (one can never know when x is present and when not)$\}$.

Why, though, should the reasonable AC theorist believe (2)? It seems that the truth of (2) would rise or fall with logical positivism, which is not obviously correct.

Or, consider the following argument against (2). Suppose, for sake of argument, that (2) is true. Thus, (2)'s truth is an ingredient of the world. Clearly, though, (2)'s truth is not an observable ingredient of the world. Hence, according to (2), one can never know the truth of (2). But all truths are knowable. Hence, since (2) cannot be known, (2) is false.

Finally, suppose that an ACE occurs. If it occurs, then it is true that it occurs. If it is true that it occurs, then someone (perhaps God) could know that it occurs. Thus, (2) is false.

Perhaps, though, Thorp means to restrict the cognizers in (2) to finite agents. Even so, it seems that if God could know that an ACE occurs, then God could tell us that it occurs. Now this need not be a necessary truth. Perhaps God could not tell us because we are unable to understand the claim. Or perhaps there are certain moral reasons that restrict God in telling us that an ACE occurs. Or perhaps the claim is about finite creatures not knowing something. After all, how do we know that God could properly tell us? Nevertheless, all we need is the possibility of God's telling us that an ACE occurs. We just need one instance. Can we be sure,

---

[63] Pacé Reid (2001).

then, that it would ever be morally permissible that God reveals an ACE to us?[64] I think that we can be sure enough. The mere possibility of God's telling us that an ACE occurs seems quite plausible—I'm content to let the strength of my argument against (2) vary in proportion to the degree of plausibility that this is a mere possibility. And if God could tell us that it occurs, then we could know that it occurs. Therefore, it seems that even we finite creatures could know that an ACE occurs. So, even the qualified version of (2) seems false. An ACE occurs only if we could know that it occurs, which is not an embarrassing result at all.

Carl Ginet expresses a complaint similar to Thorp's when entertaining the possibility of wedding agent-causation to his *simple indeterminist* theory of free action. Ginet (1997: 96) states:

> My other worry about the concept of agent-causation is that agent-causation appears to be undetectable. Given that we have a person, *S*, who is the subject of a simple mental event that has the actish phenomenal quality—a volition, say, or mentally saying a word—and this event is not causally determined by antecedent events and states, what detectable difference would it make to the situation whether the agent-causal relation between *S* and that event was present or not?

Ginet seems to be assuming that the occurrence of an ACE is not epistemically detectable. But why think this? If it occurs, then it can be known to occur. Hence, the cognitive faculties of the subject knowing that an ACE occurs would detect the ACE's occurrence.

Moreover, let's suppose that the following event is detectable: *its coming to pass that an agent is an underived source of an element intrinsic to her directly free action*. Why couldn't the AC theorist point to this event as marking the difference between a situation where an ACE is present and a situation where it is not? Presumably, Ginet would contest that this puts the cart before the horse, that the agent-causalist must provide a difference-maker more metaphysically basic than the obtaining of the agent-causal relation itself. For, Ginet (1997: 96, <u>my emphasis</u>) argues:

> What <u>independently</u> specifiable difference could distinguish between the case where the agent causal relation is present and the case where it is not? That is, what independently specifiable feature could <u>constitute</u> the agent-causal relation? … However the person and her actish mental event are realized in the ultimate constituents of reality, it is difficult to imagine what additional thing might be realized there that would *force* us to describe it as a causal relation between the person as enduring substance and the event. Whatever independently specifiable difference might be pointed to, we would be free to regard its coming about as *another event* of which the person is the subject rather than as a <u>brute</u> causal relation between the subject and the actish event.

But why must AC's being an adequate theory of free action require analyzing the primitive relation of an agent's directly producing an event into more elementary features? It seems that Ginet's demand is patently unfair to the AC proponent, as it requires something

---

[64] Tom Flint raised this question in personal conversation.

obviously incoherent. By definition, primitives are not analyzed. The AC theorist claims that the relation of direct production obtaining between an agent *qua* substance and an element intrinsic to her directly free activity is *primitive*, i.e., brute. Yet while Ginet recognizes that the AC theorist believes that this relation is *brute*, he requires that the AC theorist specify independent features that *constitute* the brute relation. It seems, then, that Ginet owes us a good argument for thinking that the agent-causalist should not take as primitive the relation of an agent's directly producing an event. But Ginet offers no such argument.

Maybe Ginet means to claim that there is no more of a reason to think that an ACE occurs than there is a reason to think that only the ACE's event-constituent (i.e., the simple mental event) occurs. Consider an ACE, (Δ•→e). Now, consider the following two propositions:

(3)   (Δ•→e) occurs.

(4)   e occurs.

So, perhaps Ginet is asking, "Given that e lies at the heart of Δ's free action, why should we think (3) is true instead of simply thinking that (4) is true? More precisely, why not deny (3) and accept only (4) as the candidate for the agent's directly free action? What evidence is there for thinking that Δ bears this special causal relation of production to e?"

This argument is a throwback to the argument from simplicity, which I discussed in detail at the beginning of this chapter. My reply is fairly simple. We should hold to (3) because (3) captures Δ's being an underived originator of an element intrinsic to her free action. (4) does not in any principled manner satisfy the condition of origination, as we saw in Chapter 5. So, since a free action intuitively requires that the agent is an underived source of some event, we should hold to (3). As we saw in Chapter 5, Ginet's view, barring an appeal to AC, is insufficient for free action. Hence, Ginet should not object to the agent-causalist's primitives on the basis of simplicity.

7.10   Conclusion

This chapter assessed many criticisms of AC, finding none of them convincing. We encountered no reason to think that one might satisfactorily account for free action without employing the notion of agent-causation. Hence, considerations from simplicity alone are of no use to our critic. We saw that Davidson's Challenge rests on a contentious proposition that the reasonable AC theorist may deny, *viz*., that an agent produces her directly free action. Confirming this result, we saw that Chisholm proposes no good reason to believe Davidson's contentious proposition. We learned that the Timing Problem is not a special problem for the AC theorist. We observed that Kane's comments—regarding our being able to paraphrase certain causal transactions into event-speak—provide little support for the categorical thesis of Eventism. We also saw that, contrary to Kane, considerations from what he calls the plurality conditions do not undermine our AC view. For, either the AC theorist can meet the plurality conditions or else the plurality conditions are orthogonal to an adequate account of a directly free action. We noted that van Inwagen provides no good reasons for denying that the AC view nonvacuously yet obviously implies that there is an undetermined free action. As we saw in Chapter 4, his argument against the possibility of an undetermined action fails. Moreover, he

offers no convincing case for thinking that the AC view is mysterious. We found that Bishop, Watson, and Kane's complaints that AC merely labels the libertarian's problem/solution are, if they are supposed to count as an objection against AC, unconvincing and unsupported. Finally, we saw little reason to think that AC should be rejected on epistemic grounds. In short, we've yet to find a

devastating argument against AC. An important question that remains is whether or not AC has the resources to explain one's freely acting for reasons. Let's turn, then, to this question.

CHAPTER 8

FREELY ACTING FOR REASONS

Free choices are reasonable and sensible in the light of a belief-desire psychology
where we choose in order to maximize the likelihood of achieving our goals.
—Richard Double (1991: 13)

8.1   Reasons-Explanation

Usually, if not always, one acts freely for reasons. In keeping with our discussion in Chapter 5, we use term "reasons" broadly, covering desires, beliefs, or some combination of these. As a working example, suppose that Sam wants some fresh air and believes that by opening the window he might satisfy his desire to have fresh air. Perhaps he also desires to stay warm and believes he might satisfy this desire by keeping the window closed. As it turns out, Sam freely decides to open the window, and his decision issues smoothly in his freely opening the window. Moreover, suppose that Sam freely opens the window *because* he desires fresh air. That is, Sam freely opens the window at least *for* the reason that he desires fresh air. Citing the reason for which one freely acts at least partly explains one's free action.[1] Any adequate theory of free will ought not gainsay this intuitive pre-philosophical datum.

Furthermore, when acting freely, there might be multiple reasons each of which favors (i.e., provides a rationale for) the same action. For example, suppose that Sam also desires that a bug be out of the room and believes that by opening the window he might satisfy this desire. However, even though Sam has two desires each of which favors opening the window, we may consistently suppose that Sam freely opens the window for the reason that he desires fresh air but *not* for the reason that he desires that the bug be out of the room.

Cases like this are no surprise to many philosophers. Donald Davidson (1963: 9) observes that "a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it." Carl Ginet (2002: 387) nicely expresses the point:

> There are two essential aspects to a reasons explanation: (1) the set of the agent's propositional attitudes offered as explanans must constitute a rationale for acting in the way the agent acted, and, *in addition*, (2) that rationale must be the reason (or one of the reasons) for which the agent acted in that way; and the first does not entail the second.

It would seem, then, that an adequate account of free will (i.e., an account that renders free will coherent and intelligible) should have resources for distinguishing between acting freely

---

[1] I assume the synonymy of the following locutions: "subject S acts (at least partly) for the reason that she has desire D," "S acts (at least partly) for desire D," "S acts (at least partly) on desire D," "S acts (at least partly) from desire D," "S acts (at least partly) because she has desire D," "D is part of the explanation of S's acting," and "S acts in order to satisfy her desire D." In the sequel, I drop but still assume the qualification "(at least partly)".

for only one of a few reasons, even if more than one of these few reasons point to, teleologically favor, or could provide a rationale for the very same action. Providing a reasons-explanation for an undetermined free action goes a long way toward answering the Intelligibility Question: can we make sense of a free will—a free will worth having—that requires indeterminism? I shall argue that the AC theorist can render intelligible a rational free action, for she can give sufficient and informative conditions for one's acting freely for a reason.

A dominant position among contemporary philosophers has it that a reason explains an action if and only if it (together with other environmental factors) produces the action in the right sort of way (e.g., non-deviantly). Call this position *causalism* and its adherents, *causalists*. Causalists have a straightforward way of unpacking our target distinction. They account for *Sam's freely opening the window for the reason that he desires fresh air* by claiming that Sam's desire for fresh air produces his freely opening the window. Likewise, though Sam freely opens the window, he does not do so for the reason that he desires that the bug to be out of the room, when the latter desire does not produce his freely opening the window.

If causalism is right, then our version of AC cannot account for an agent's performing a directly free action for a reason. Why? Because our AC view implies that an agent-causal event (ACE), *agent Δ's directly producing a certain intention*, just is the agent's directly free action. Moreover, our view implies that nothing can produce an ACE. *A fortiori* the agent's reasons cannot produce his directly free action. Hence, if Δ freely acts for a reason only if this reason produces her action, as causalists require, then our AC view implies that no one can perform a directly free action for a reason.

Not surprisingly, Timothy O'Connor (2000a: 85) maintains that causalism is wrong, arguing that his agency theory "permits an understanding of how reasons might explain an action without directly causing it." He attempts to give conditions consistent with his agency theory that suffice for an agent freely acting for a reason. Some thinkers remain unconvinced. Richard Feldman and Andrei Buckareff (F&B) argue that O'Connor's conditions are insufficient.[2]

In what follows, I briefly outline O'Connor's conditions for a reasons-explanation of a free action and explain F&B's alleged counterexample. While I argue that F&B's criticism is not telling, I suggest that one could strengthen the defensibility of O'Connor's account by using my account of causal contribution. I entertain and dismiss an objection against grafting my account of causal contribution onto O'Connor's theory of AC. If I'm right, the AC theorist has taken great strides in making sense of an undetermined and rational free action.

This chapter, then, addresses whether the agent-causalist can make sense of an agent's freely acting for a reason. §8.2 outlines O'Connor's attempt to provide sufficient, consistent, and informative conditions for an agent's rational free action. §8.3 outlines F&B's criticism of O'Connor's attempt. §8.4 defends O'Connor's account, providing a reasoned evaluation of F&B's case. I go on in §8.5 to argue that even if we concede for sake of argument that F&B's criticism sticks, one could modify O'Connor's account by wedding it to my account of causal contribution. §8.6 closes this chapter by examining an objection that one may be tempted to advance against one's using my account of causal contribution to deepen a theory of reasons-explanation.

---

[2] Feldman & Buckareff (2003). Randolph Clarke (2003: 138-144) criticizes O'Connor in much the same way.

8.2   Tim O'Connor's Account

Recall that causalists hold that a reason explains an action if and only if the reason is an essential part of what produces the action. The AC proponent may consistently hold that the causalist offers a sufficient but unnecessary condition for acting on a reason. The AC theorist denies that the causalist can offer a sufficient condition for *freely* acting on a reason; hence, the causalist's condition is unnecessary. So if the causalist cannot provide sufficient, consistent, and informative conditions for an agent's freely acting on a reason, can agent-causalists do any better? O'Connor (2000a: 85) thinks so, saying, "There is more than one way that reasons could figure in the explanation of action. The agency theorist needs only to give schematic *sufficient* conditions, conditions consistent with the agency theory's conception of *free* action."[3] Let's investigate how O'Connor's attempt survives a recent criticism.

Recall that on our AC theory, an ACE is the agent's directly producing a certain (state of) intention. Intentions, like beliefs or desires or hopes or wishes etc., are propositional attitudes. A propositional attitude is, aptly enough, an attitude toward a proposition. The proposition is said to be the content of the attitude.

For example, Fipe has a belief. He believes that he is failing logic. He has an attitude, an attitude that is a believing/assenting. Fipe's attitude is toward a proposition, the proposition that he is failing logic. The proposition *that he is failing logic* is the content of Fipe's belief. Consider the propositional attitude of desiring. Fipe desires to pass his logic exam. That is, Fipe desires that he passes his logic exam. His desire is an attitude toward the proposition that he passes his logic exam. Similarly, Fipe has an intention. He has an intention to study for his upcoming logic exam. That is, Fipe intends that he studies for his upcoming logic exam.

Propositional attitudes may be more or less complex, depending on the richness of their contents. For example, the content of Fipe's belief that he is a rock-climber who failed his logic class is much richer than the content of Fipe's belief that he is a rock-climber. Intentions likewise may be distinguished by differences in their contents. Notice the difference between the following intentions. Fipe intends to finish college. Fipe intends to keep the promise he made to his mother by finishing college in five years.

Now, crucial to O'Connor's theory is the idea that agents directly produce (i.e., agent-cause) intentions that have a content rich enough to represent the desire on which the agent putatively acts. An intention's content clearly may involve, include, or make reference to a desire. For example, Fipe intends to pacify his desire for a noontime milkshake by skipping class. That is, Fipe intends to skip class for the sake of pacifying his desire for a noontime milkshake. Skipping class may even happen to satisfy another one of Fipe's desires, even when Fipe does not intend to appease his other desire.[4]

Given this brief prolegomenon on how roughly to understand the contents of propositional attitudes such as beliefs, desires, and intentions, we turn to O'Connor's basic account of reasons-explanation for free action. O'Connor (2000a: 86) explicitly submits the

---

[3] O'Connor's conditions are designed to suffice for explaining a person's free action in terms of an antecedent desire. O'Connor (2000a: 86) claims that his account is schematic in the sense that with minor alterations one could explain a person's action in terms of her prior intention, or in terms of a prior belief, et. al.

[4] To say that one does *not intend* to x is not to say that one *intends not* to x.

following conditions, which he maintains are sufficient for at least partly explaining an agent's free action in terms of her prior desire:

The agent [freely] acted then in order to satisfy his antecedent desire that Θ, if

(1)   prior to this action, the agent had a desire that Θ and believed that by so acting he would satisfy (or contribute to satisfying) that desire;

(2)   the agent's action was initiated (in part) by his own self-determining causal activity, the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-act-here-and-now-so-as-to-satisfy-Θ;[5]

(3)   concurrent with this action, he continued to desire that Θ and intended of this action that it satisfy (or contribute to satisfying) that desire; and

(4)   the concurrent intention was a direct causal consequence (intuitively, a continuation) of the action-triggering intention brought about by the agent, and it [non-deviantly] causally sustained the completion of the action.

Condition (1) is rather uncontroversial, as it simply ensures that we have a typical case of an agent having a reason that could count as a candidate for the reason on which the agent acts freely. Recall that prior to Sam's freely opening the window, Sam desires fresh air and believes that by opening the window he would satisfy that desire. Hence, so far we may use Sam's situation to illuminate O'Connor's sufficient conditions, as Sam's case meets condition (1). So far so good.

Since O'Connor aims to give sufficient conditions for a reasons-explanation of a free action *qua agent-causalist*, at least one of his conditions should imply that an ACE occurs. Such is the role of (2). The self-determining causal activity mentioned in (2) just is an ACE. It is the agent's directly producing an intention. O'Connor calls this ACE the agent's "decision."

But (2) goes beyond simply stating that an ACE occurs. (2) also qualifies the event-constituent of the ACE. Again, this event-constituent is an intention directly produced by the agent. (2) implies that the intention has quite a bit of content. It is an intention not only to act but also to act in a certain way. As we've described Sam's case, the act is opening the window. So, with (2) in mind, we may imagine Sam intending to open the window. Furthermore, (2) prescribes a representational content richer than simply intending to open the window. In addition, the agent's intention includes performing a specific action *in order to* satisfy a desire that he possesses. Applying this to Sam's case, we imagine that in addition to Sam's desiring fresh air and believing that his opening the window might satisfy this desire, Sam also intends to open the window in order to satisfy his desire for fresh air. It seems that according to O'Connor, Sam's intention links a particular desire (*viz.*, wanting fresh air) with a particular action (opening

---

[5] It's hard to see whether O'Connor uses "Θ" to represent the desire or the content of the desire. I'll assume that Θ is the desire, even though O'Connor uses the locution "the desire that Θ". Nothing of substance should hinge on this assumption.

the window). Again, we may continue to use Sam's situation to illustrate how O'Connor's conditions are supposed to be sufficient for a reasons-explanation of a free action.

Conditions (3) and (4) aim to ensure that the intention just referenced in (2) is not undercut or short-circuited in some way, thereby making it explanatorily irrelevant. Consider a case that (3) rules out. Suppose that immediately after Sam forms the intention to open the window in order to satisfy his wanting fresh air, Sam no longer wants fresh air and no longer has any intention to open the window. Then, even if he somehow opens the window, this action will not clearly satisfy his desire for fresh air, since he doesn't even have this desire anymore. He completes an action that he no longer has any intention of performing. Neither his desire nor his intention is there to be satisfied. (3) is surely consistent with (1) and (2), and since O'Connor only needs sufficient conditions, (3) should only help. Finally, condition (4) is designed to preclude deviant or wayward causal chains.[6]

The upshot of O'Connor's account is that an agent's free action begins with and is nondeviantly causally sustained by an intention. The content of this intention is to act in a precise way in order to satisfy a precise desire, where he has this desire both before the action and throughout the duration of the action. The agent, then, produces intentions that successfully refer to antecedent and concurrent desires.

In Sam's case, his intention to open the window in order to satisfy his desire for fresh air continues (or issues in an intention with the same content that continues) until Sam completes his action. Sam's initial agent-caused intention and the concurrent intention (intuitively, a continuation of the agent-caused intention) successfully refer all the while to his desire for fresh air. Moreover, Sam's intention, referring to this particular desire, intuitively suffices for Sam's aiming all the while to complete his act of opening the window. He intends to open the window. He intends to satisfy a particular desire. Sam completes his action, thereby succeeding in doing what he intends to do. Sam's action satisfies the very desire that he intends to satisfy.

By O'Connor's lights, this suffices for Sam's freely opening the window for the reason that he wants fresh air. Intuitively, Sam's desire is at least part of the explanation of Sam's freely opening the window. Thus, it is plausible that the AC proponent can provide consistent conditions sufficient for acting freely for a reason.

8.3   On Feldman & Buckareff's Putative Counterexample

F&B (2003: 135) claim that O'Connor is a "pure agency theorist." Pure agency theorists hold that "reasons are not even partial causes of the [free] behavior for which they are reasons." F&B state their thesis as follows:

> Denying that reasons are even partial causes of actions seems to open the pure agency theory up to a potentially serious objection. It seems that the most it can say about the reasons for which an agent freely acts is that they are present in the agent before and at the time the agent causes the intention that results in behavior.

---

[6] The problem of deviant causal chains is a problem for action theorists in much the same way as the Gettier problem is a difficulty for epistemologists. Regarding the former problem, see my comments in Chapter 5, §5.1. Since deviant causal chains is not a special problem for the AC theorist, I assume that there is an adequate solution.

However, <u>there is good reason to believe</u> that it is not true that an agent did A for reason R if S had a reason R, where R had no causal role, and S caused A.[7]

So F&B attempt to give a counterexample against O'Connor's claim that for some agent S and some desire Θ and some action A, O'Connor's conditions require that Θ is part of the explanation of S's freely doing A.

F&B (2003: 138-9) have us imagine the following case. Agent S has two desires, Θ and Φ, each of which would be satisfied by action A. Without altering the neural realization of Θ, a neuroscientist ensures that Θ is incapable of having any causal effect on the initiation of S's action. Θ is completely causally impotent.[8] Φ, on the other hand, figures normally in the etiology of S's action.

F&B hold that it is plausible that "Φ plays an explanatory role in understanding the initiation of S's action while Θ does not." They argue:

> If this is correct, then O'Connor's conditions can be satisfied even though Θ does not in fact explain S's action. For O'Connor's conditions to be satisfied <u>in this example</u>, it is required that the person agent-cause an intention to satisfy the causally blocked desire. Thus, although Θ itself is causally ineffective, S does form an intention to act in order to satisfy Θ. We think that in such a case, Θ is not actually a reason for which S acted.[9]

Let's use our working example of Sam to illustrate F&B's point. Let Θ be Sam's desire for fresh air, and let Φ be Sam's desire that a bug be out of the room. Aptly enough, agent S is Sam. As F&B set up their example, the neuroscientist renders Sam's desire for fresh air wholly causally inefficacious, but Sam's desire that the bug be out of the room figures normally in the etiology of Sam's opening the window. Sam *qua* enduring substance directly produces an intention to open the window in order to satisfy his desire for fresh air. This intention plays the right sort of causal role vis-à-vis the completion of Sam's opening the window.

So, F&B maintain that since Θ is causally inert, Sam does not act for reason Θ, even though the intention to satisfy Θ has causal efficacy. F&B insist that Sam's intention, which successfully refers to Θ and is the event component of Sam's decision, does all of the relevant explanatory work. Since Sam's Θ causes nothing, it plays no explanatory role. Thus, F&B conclude that Sam's desire, Θ, is not even part of the explanation of Sam's freely opening the window. Therefore, O'Connor's conditions are insufficient.[10]

---

[7] F&B (2003: 135-6, <u>my emphasis</u>).

[8] That is, completely causally impotent with respect to the agent's action. Θ is not part of the etiology of the agent's action.

[9] F&B (2003: 138, <u>my emphasis</u>).

[10] Clarke's (2003: 139) crucial premise can be distilled as follows. Necessarily, if (i) O'Connor's four conditions hold, and (ii) the agent's desire that Θ causes neither the action-triggering intention nor the concurrent intention nor any part of the action downstream of the action-triggering intention, and (iii) for some desire Φ different than Θ, either Φ causes the action-triggering intention or Φ causes the

8.4   Assessing Feldman & Buckareff's Criticism
          There is a small worry concerning how F&B set up their example. They claim to satisfy
O'Connor's conditions, yet they say that Φ figures normally in etiology of S's action, thereby
playing an explanatory role in understanding the initiation of S's action. As we've seen from
previous chapters, there are two ways that Φ might figure in the etiology of an event. Either Φ
produces the effected event or else Φ merely-contributes causally to the effected event.
          If by "Φ figures normally in the etiology of S's action" F&B mean Φ produces S's action,
then F&B do not satisfy O'Connor's conditions. For, S's free action begins with S's ACE,
which in this example is *S's directly producing an intention to A in order to satisfy Θ.* Since
nothing can produce an ACE, Φ cannot produce an ACE. F&B do not attempt to object to
O'Connor's argument for thinking that nothing can produce an ACE. So, it seems that the only
way Φ can figure in the etiology of S's action is by merely contributing causally to the ACE.
For sake of argument, let's assume that Φ merely-contributes causally to S's ACE.
          Now, I think that one should agree with F&B that, plausibly, a complete explanation of
S's action includes the etiology of the action. And since by our assumption this etiology
includes Φ, Φ is part of an explanation of S's action. Nonetheless, it is worth noting that it is not
yet obvious (to me, at any rate) that Φ is a reason *for which* S performs A. It is still unclear that
S performs A *for* reason Φ. So far as the example is concerned, S does not even intend to satisfy
Φ.
          The more important issue, though, concerns the explanatory status of Θ. Recall F&B's
(2003: 136) thesis, "[T]here is good reason to believe that it is not true that an agent did A for
reason R if S had a reason R, where R had no causal role, and S caused A." But have F&B given
this good reason? I think not, and the balance of this section explains why. Moreover, even if
I'm wrong about this and F&B force O'Connor to retrench, my next section, §8.5, provides a
strategy on behalf of the AC theorist for answering F&B's criticism on its own terms.
          So, what is the argument F&B advance for thinking that S does not act for reason Θ?
Unfortunately, they offer no argument. Rather, after proffering their thought-experiment, they
(2003: 138) conclude, "We think that in such a case [i.e., in F&B's example], Θ is not actually a
reason for which S acted." They (2003: 139) stress, "Θ itself is causally cut off from the
intention to satisfy Θ."
          But how is this a good reason for thinking that Sam does not act on Θ? How does this
show that Θ does not even partly explain why Sam freely opens the window? There seems to be
no argument here but just a bare statement of faith. Their thought-experiment is not a clear
counterexample against the sufficiency of O'Connor's conditions. I see no good reason for
thinking that one acts freely for a reason only if the reason causes the action.
          Interestingly, F&B (2003: 139) predict this sort of response, claiming:

> We realize that O'Connor may contend that in such a case Θ really is a reason for
> the action, and that Θ does partially explain S's behavior. This strikes us as
> incorrect. Θ was inert. At most the intention to satisfy Θ played a role. The
> reasons that favor this conclusion go beyond a mere assertion on our part that
> reasons [for which one freely acts] are causes.

concurrent intention, then it is not the case that the agent freely acts in order to satisfy Θ. The points I
make in response to F&B apply equally well against Clarke's premise.

Again, thus far, the only reason for dismissing O'Connor's account is that Θ is causally inert. However, the lattermost claim of this quotation suggests that F&B have reasons in addition to their assumption that Θ cannot explain if causally inert. Evidently, F&B (2003: 139-40) attempt to buttress their thought-experiment with additional considerations. What are these considerations and do they count as a cogent argument for thinking O'Connor's conditions are insufficient?

F&B (2003: 139) continue by rightly noticing that in formulating his theory, O'Connor "switches from talk of reasons explanations to talk of when an 'agent acted in order to satisfy his antecedent desire.'" F&B then ask us to compare:

(A)   Desire D is part of the explanation of S's freely doing A

(B)   S freely did A in order to satisfy desire D

Now F&B grant that O'Connor's conditions suffice for (B) but suggest that O'Connor should instead be giving conditions sufficient for (A). F&B claim:

> [In the case where (A) and (B) are not equivalent], we take (B) to describe what S had in mind in doing A. And we grant that our example is not a counterexample to the proposal, since S did have both desires [Θ and Φ] in mind in performing the action. But then our example is one in which (B) is true but (A) is false.[11]

Unfortunately, with the exception that Θ is causally inert with respect to the initiation of S's behavior, F&B offer no reason for thinking that their thought-experiment implies a false instance of (A). Two important points follow.

First, it follows that F&B offer no reason for thinking that their example is a counterexample against O'Connor's conditions sufficing for (A). Again, the only reason for thinking that their example implies that (A) is false is that Θ is causally impotent. But why should this be a reason? F&B seem to assume without argument the principle that (A) requires that D plays a causal role in S's freely doing A. That is, they assume

(A*)   Desire D is part of the explanation of S's freely doing A *only if* D is part of the etiology of S's freely doing A.

But why think (A*) is true? At first glance, it seems to be blind faith in causalism.

Second, it follows that F&B offer no reason for thinking that their example shows that (B) is insufficient for (A). In other words, for all F&B have shown, (B) may suffice for (A).

F&B admit that (B) describes what S has in mind in doing A. They correctly note that simply having a desire in mind when acting is not sufficient for acting on that desire. However, F&B offer no argument for thinking that (B) *only* describes what S has in mind in doing A. Recall their (2003: 135-6, <u>my emphasis</u>) thesis, "It seems that <u>the most</u> [that the pure agency theorist] can say about the reasons on which an agent freely acts is that they are present in the

---

[11] F&B (2003: 139-40).

agent before and at the time the agent causes the intention that results in behavior." Unfortunately, this thesis does not follow from the fact that (B) describes what S has in mind in doing A, and F&B advance no other premises to earn this inferred thesis.

Moreover, (B) *does* describe more than just what S has in mind in doing A. Notice that (B) has an intentional component. (B) does not only imply that S freely acts while believing (truly) that he has desire D. Rather, the agent *aims* to satisfy his desire. Again, in our working example, Sam directly produces an intention that successfully refers to Θ. Throughout the performance of his action, Sam aims to complete his act of opening the window. He intends to open the window. He intends to satisfy Θ. Sam completes his action, thereby succeeding in doing what he intends to do. Sam's action satisfies the very desire that he intends to satisfy.

Here's another reason to think that (B) describes more than just what S has in mind while acting. Suppose that Sam, throughout the performance of opening the window, believes truly that he has desire Φ, where Φ is his desire that the bug be out of the room. If (B) only describes what Sam has in mind during his action, then Sam freely opens the window in order to satisfy Φ. But this contradicts the plausible claim that Sam does not freely open the window in order to satisfy Φ. Sure, we've admitted that Φ is part of the explanation of Sam's freely opening the window, but this does not clearly imply that Sam acts *in order to* satisfy Φ. He does not even aim to satisfy Φ. If this is right, then we have a case where (A) is true while (B) is false. Nonetheless, it may still be that (B) implies (A).

It seems to me that (B) implies (A). Suppose, for example, that F&B's thought-experiment is self-consistent. We may then consistently imagine it fleshed out as follows. Sam's friend, Geoffrey, asks Sam to offer part of an explanation as to why he opened the window. Sam responds, "I wanted fresh air." Geoffrey educates Sam, telling him that his wanting fresh air in no way caused him to open the window. Sam replies:

> Oh, that doesn't matter. I initiated my own action, and I acted because I wanted fresh air. After all, I intended to get some fresh air by opening the window, and I did just that. And guess what? I also wanted to get a better view of the mountains and those thick windowpanes were obstructing my view. While I acted, I believed that I wanted a better view of the mountains. But I did not open the window in order to satisfy my desire for a better view. That desire was unintentionally satisfied—a foreseen but unintended bonus, if you will. Rather, I freely opened the window in order to get some fresh air. It was really stuffy in here, wouldn't you agree?

Sam continues:

> Oh, by the way, I already met that neuroscientist. She came in here after I opened the window and told me what she had done. She also said that she made my desire for a better view completely causally inefficacious. I'm inclined to think that my desire for a better view in no way explained my opening the window. After all, it didn't even contribute causally to my action, and I did not even intend to act on it. Wow, I had no idea that neuroscientists could do that sort of thing…

F&B provide no reason for holding that Sam speaks falsely. It seems to me that Sam's desire for fresh air, Θ, is part of an explanation for his freely opening the window. That is, it

seems that (A) is true even in Sam's case, and Sam's case meets F&B's restrictions. It seems that O'Connor may consistently say that even though Φ explains Sam's action insofar as it is a causal contributor to his action, Θ is also a reason for which Sam acts even though Θ has no causal effect on the initiation of Sam's action.

8.5 Adjusting O'Connor's Basic Account

For sake of argument, suppose that F&B's criticism of O'Connor's account of reasons-explanation sticks. I shall now argue that one can meet F&B's objection on its own terms by appealing to the notion of causal contribution. The strategy is simple. We weld to O'Connor's conditions a consistent clause requiring that the putative reason for which the agent acts contributes causally to the agent's action. I'm very much inclined to think that this amended account is in keeping with what O'Connor has always suggested even if he has not always been entirely explicit.

This section begins with an attempt to make explicit O'Connor's commitments to an agent's reason's contributing causally to her free action. Even if I am wrong about what O'Connor's precise commitments really are, our discussion should motivate the claim that an agent-causalist may accept that reasons can merely-contribute causally to free actions. Moreover, if this is right, then the AC theorist should at least be able to provide sufficient conditions for a reason's being part of the explanation for an agent's free action.

8.5.1 Does O'Connor Reject a Reason's Contributing Causally to a Free Action?

Recall that F&B (2003: 135) call O'Connor a pure agency theorist and that they define a pure agency theorist as someone contending "that reasons are not even partial causes of the [free] behavior for which they are reasons." At first glance, it would seem that a causal contributor just is a partial cause. So, if F&B are right that O'Connor maintains that reasons are not even partial causes of one's free action, then it would seem that O'Connor cannot hold that reasons contribute causally to one's free action.

Unfortunately, F&B use of the locution 'partial causes' here is misleading. They do not tell us what it means to be a partial cause. Instead, they give us two examples of agency theorists who contend that reasons are partial causes. (Actually, they note only one agency theorist, Randolph Clarke, but he has held two distinct views.) First, Clarke (1993) holds that an agent qua substance produces the complex event that is her reason's producing her free action. Second, turning a new leaf, Clarke (1996a) holds that an agent qua substance and the agent's reasons jointly produce the agent's free action. Here, the agent and her reasons are co-producers of the action.

These two examples are consistent with the claim that some C is a partial cause of E only if C produces E. More precisely, if one assumes that the examples exhaust the ways that a reason could be a partial cause, then the examples imply that some C is a partial cause of E only if there is some C* such that C*≠C and C* produces E at the same time that C produces E. If this is right, then O'Connor is certainly a pure agency theorist. For, he holds that nothing can produce an ACE, which together with the assumption that C is a partial cause of an event only if C produces the event implies that there can be no partial cause of an ACE, including reasons. This reading of 'partial cause' confirms F&B's claim that O'Connor is a pure agency theorist.

Nonetheless, there is another reading of 'partial cause' that perhaps better squares with our natural language and pre-reflective beliefs. On this reading, C is a partial cause of E *iff* C contributes causally to E. If this is right, then O'Connor need not be a pure agency theorist, for

197

he may consistently believe that reasons do not produce but still merely-contribute causally to an agent's free behavior for which they are reasons. If so, we should be able to add such a causal-contribution-clause to O'Connor's conditions, thereby rendering F&B's putative counterexample a non-starter.

Before adjusting O'Connor's explicit account, let's first see some textual evidence that suggests that O'Connor holds that reasons may contribute causally to an agent's free action. In the introductory chapter of his (2000a) book *Persons and Causes*, O'Connor describes how he will advance a basic account of reasons-explanation for a free action—this basic account simply consists of the four conditions that I just defended from F&B's criticism. O'Connor's (2000a: xiv-xv) then states in his introduction:

> Near the end of the chapter, I suggest a refinement of the basic account in which the having of reasons generates or raises a *carried tendency* to act in particular ways, which tendency probabilistically structures the basic agent-causal activity. It remains up to the agent, nonetheless, to determine which such tendency will be acted on.[12]

This quotation suggests that there is more to O'Connor's account of reasons-explanation than the conditions that F&B criticize.

Later, in his third chapter, when discussing the AC theories of Richard Taylor and Roderick Chisholm, O'Connor (2000a: 52 note 22) again anticipates his account of reasons *causally* explaining one's free action, saying,

> In chapter 5, I take up the matter of causal influence (in particular, of an agent's reasons) over an agent's free activity, suggesting how one might understand this in a way that does not involve factors that produce it.

Notice that O'Connor here describes the influence of reasons on an agent's free activity as being *causal* influence. This further suggests that O'Connor's basic account is not fully developed, since the basic account in no way implies that reasons play any causal role on free action—as F&B bring into stark relief.

Just five pages later, O'Connor (2000a: 59 note 34) says:

> [T]here will be a large number of necessary causal conditions for the capacity to exercise such control [i.e., for the capacity to agent-cause an event]. As promised earlier, I try to account for the presence of such enabling (as well as constraining and influencing) factors in a way that is consistent with the thesis that agent-causal events are not causally produced.

An agent's directly producing an event (i.e., an ACE) strictly implies a capacity to directly produce an event. And O'Connor claims that this capacity has necessary causal conditions. Hence, since necessary causal conditions are causal contributors, it follows that there may be

---

[12] I'm not sure what it means to act on a tendency. One may act on a belief, on a desire, on a reason, on a previous intention, etc. Perhaps O'Connor means that one acts on a reason *that* generates a tendency to act.

causal contributors of an ACE. Moreover, it may be that an agent's reasons are among the mere-causal-contributors of her free action. Chisholm (1978: 630) expresses this point nicely:

> And still another point about the causal efficacy of one's reasons. If what I do does not have a *sufficient* causal condition [e.g., if nothing produces what I do], it will all the same have indefinitely many *necessary* causal conditions— indefinitely many conditions which are such that, if they hadn't occurred, I wouldn't have done what I did do. My reasons *can* be an essential part of these. For there are many things I undertake which are such that it would not be within my power to undertake them unless I had a reason or motive for undertaking them.

Nonetheless, no causal contributor of an ACE produces the ACE. They *merely*-contribute causally to the ACE.

In O'Connor's chapter outlining the signature-features of his own AC theory, O'Connor (2000a: 73) states:

> In the next chapter I give a general analysis of reasons explanations of free actions in noncausal terms. (However, there are some contingent features of human agents that indicate that the exercise of active power has causal structure. This is taken up in section 5.5 (sic, 5.4).)

Context makes it clear that O'Connor's use of 'causal structure,' here, does not refer to an ACE's having a causally complex structure. Rather, O'Connor is claiming that there is a causal structure between an ACE and prior events.

Yet again, a page later, O'Connor (2000a: 74-5 note 14, *my emphasis*) stresses:

> In the next chapter, I concede that facts about ordinary [free] choices by human beings can be understood *only by ascribing a causal role to their reasons*. In doing so, however, I suggest that there is a way to understand the causal role to reasons apart from their producing agent-causal events (which in any case is impossible).

This quotation is most damaging to F&B's attempt to provide a counterexample to O'Connor's account by stipulating that the agent's reason on which he freely acts plays no causal role whatsoever in the agent's free behavior. Recall that in their example a neuroscientist renders S's desire Θ causally impotent. Of course, F&B rightly note that O'Connor's basic account does not commit him to thinking of reasons as playing a causal role on free action. But given that O'Connor forewarns his reader that he'll develop his basic account in greater detail, one should charitably take his basic account as a first approximation.

In at least two other essays predating F&B (2003), O'Connor summarizes the simple point that an agent's reasons may causally influence her free action. For example, O'Connor (2000b: 117, *my emphasis*) says:

> There are some contingent features of human agents that indicate that *the exercise of active power has further causal structure in the way reasons govern it.*

Reasons move us to act, and some do so much more strongly than others. In ch.5 of my book, I propose that we think of the agent's states of having reasons to act in various ways as structuring the agent causal capacity, such that the agent's freely choosing an action type will have some objective tendency to occur, one which fluctuates over time. Even if this is accepted, it remains true that, in contrast to mechanistic causation [i.e., causal production], it remains up to the agent to decide how to act. The tendency-conferring state of having a reason *does not itself generate the action*; it disposes the agent himself to initiate an action sequence.[13]

And elsewhere, he (2002b: 352) explains that in his (1995b) essay 'Agent Causation' he gave a "very attenuated role" to reasons "in which my now having a reason to A is a necessary causal condition on my now causing the intention to A." Attenuated as it is, it is still causal contribution. O'Connor (2002b: 353, *my emphasis*) continues, saying that when acting freely,

I am the sole causal factor directly generating my intention to A (not a co-cause along with my reasons, *as in* Clarke's view), *but my doing what I do is shaped, causally, by my total motivational state*.

Here O'Connor contrasts his view of reasons merely contributing causally to a free action with Clarke's view, which entails that reasons produce the free action.

8.5.2   Adding to O'Connor's Account

O'Connor's lack of maximal clarity is partly to blame for why someone might ignore any further amendment to his basic account. For example, O'Connor (2000a: 86) falls prey to oversight when introducing the first approximation of his account, stating, "[Conditions (1)-(4)], I maintain, are enough to explain an action in terms of an antecedent desire." This isn't quite right, since if they were enough then they would not need any further development. So perhaps this is what misleads F&B. A careful and generous reading of his text, however, suggests that one should take O'Connor's claim as tentative. For, as we've seen, O'Connor clearly warns his reader in advance that these four conditions do not account for the causal role of reasons, and he promises to refine his theory accordingly so that it does provide a causal role to reasons.

I recommend that we add a fifth condition to O'Connor's four—*viz.*, Θ merely-contributes-causally to the agent's self-determining activity (ACE). The finished account, then, reads:

The agent freely acted then in order to satisfy his antecedent desire that Θ, if

(1)   prior to this action, the agent had a desire Θ and believed that by so acting he would satisfy (or contribute to satisfying) that desire;

---

[13] O'Connor (2000b: 117, *my emphasis*).

(2)   the agent's action was initiated (in part) by his own self-determining causal activity, the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-act-here-and-now-so-as-to-satisfy-Θ;

(3)   concurrent with this action, he continued to have desire Θ and intended of this action that it satisfy (or contribute to satisfying) Θ;

(4)   the concurrent intention was a direct causal consequence (intuitively, a continuation) of the action-triggering intention brought about by the agent, and it [non-deviantly] causally sustained the completion of the action; and

(5) Θ merely-contributed-causally to his self-determining activity.[14]

8.5.3   How Does Θ Contribute Causally to the ACE?

Now when O'Connor (2000a: 95-99) develops his account of reasons-explanation in more detail, he aims primarily to shed light on the notion of how reasons could influence a free agent to greater or lesser degrees. He attempts to characterize how some reasons influence us more weightily or heavily than others. After all, reasons incline us to act, and we apparently *feel* some reasons inclining us more strongly than others do.

For example, the enemy captures a soldier and demands that he freely disclose military secrets. Though it is in his power to remain quiet freely, the soldier timidly but freely gives up the secrets for whatever reasons incline him to give up the secrets. In a nearby world, suppose instead that the soldier freely remains quiet. The enemy straps the soldier to a torture rack, again demanding that he freely give up the secrets. The soldier feels more inclined to tell all, but he freely keeps quiet. He undergoes mild torture. Still, he remains quiet freely. The level of torture increases but nonetheless leaves room for him to act freely. Imagine that he freely gives up the secrets for reasons that incline him to give up the secrets. Now, compare the two worlds. In each world,
the soldier freely tells the enemy the secrets. Intuitively, the tortured soldier feels more inclined to give up the secrets than the soldier acting timidly does. The tortured soldier's reasons more weightily influenced him to give up the secrets freely.

In an attempt to account for the degree of influence of reasons on free action, O'Connor speaks of the agent having objective propensities or objective tendencies to act in certain ways. Here are two texts that I think best capture O'Connor's view of the degree of influence of reasons on free action.

> [R]ecognizing a reason to act induces or elevates an objective propensity of the agent to initiate the behavior. … Again, my reasons structure my activity, *not just in the rough manner of partitioning the possible outcomes into those comparatively few that are genuinely available and the many others that are not*, but also in the more fine-grained manner of giving me, qua active cause, relative

---

[14] See APPENDIX 1 for my account of one event's contributing causally to another event.

tendencies to act. These are tendencies that it remains entirely up to me to act on or not; what I do is not *simply* the consequences of the vagaries of 'chancelike' indeterministic activity such as may be true of quantum phenomena.[15]

I suggest that our beliefs, desires, and general intentions—in short, our reasons— contribute to dynamical propensities in us to cause an appropriate intention to act to occur. Prior to acting, there is an objective propensity for me to cause an intention to A, a propensity for me to cause an intention to B, and so on, for however many options I have some positive inclination to pursue. But none of these states are (individually or jointly) indeterministic causes that 'trigger'— directly bring about—my causing (say) an intention to A.[16]

Now these texts are deep, murky, and difficult to understand.

Nonetheless, the AC theorist need not have an account of a reason's degree of influence on free action in order to have an account of a reason's merely-contributing-causally to the free action. And the claim that a reason may contribute causally to one's free action is enough for the AC theorist to defeat F&B's criticism.

8.5.4   How the New Account Handles Feldman & Buckareff's Objection

I have offered my own account of event causal contribution.[17] Given my account, what would it look like for desire Θ to merely-contribute causally to the agent's self-determining activity? First, the agent's self-determining activity just is an ACE. The crux, then, centers on how Θ might merely-contribute to an ACE.

My account of causal contribution, to review, modifies Chisholm's 1986-account. Painting in broad strokes, there are two ways an event C might contribute to another event E. First, there is a sufficient causal condition of E and C is related in the right sort of way to this sufficient causal condition. For example, C is a part of a minimal sufficient causal condition of E. Nevertheless, this first way C could contribute to E does not apply to the present discussion, where desire Θ contributes to an agent's ACE. For, since nothing could produce an ACE, there will be no sufficient causal condition of the ACE. Second, E has a necessary causal condition and C is related in the right sort of way to this necessary causal condition. Suppose, for example, that C just is a necessary causal condition of E. Intuitively, C enables E. Given the definition of a necessary causal condition, it follows that E occurs and the natural laws imply that E occurs only if C occurs at or before the time of E. Equivalently, *C's failing to occur* is a sufficient causal condition of *E's failing to occur*.

So, how could desire Θ merely-contribute to agent S's ACE? There are a few ways it could go. Consider the following limiting case. S's desire Θ is a direct necessary causal

---

[15] O'Connor (2000a: 97, *my emphasis*).

[16] O'Connor (2002a: 136-7).

[17] For my account, see Chapter 6 or APPENDIX 1.

condition of his ACE.[18] Here, Θ enables the ACE straightaway. The natural laws imply that if S did not have Θ, then S's ACE does not occur. The laws imply that S performs this directly free action only if he has desire Θ. Clearly Θ contributes to S's directly free action. However, nothing—not even Θ—produces S's directly free action. The other ways Θ might merely-contribute get more sophisticated, but they still fall under my account of causal contribution.[19]

This limiting case suffices for showing how F&B's criticism fails to undermine our modified account of reasons-explanation, i.e., O'Connor's basic account plus condition (5). Let's return to F&B's putative counterexample. As F&B describe their case, agent S has a desire Θ, S performs a directly free action (*viz.*, S directly produces an intention to act so as to satisfy Θ), yet Θ is in no way part of the etiology of S's free action. A neuroscientist ensures that Θ is incapable of having any causal effect on the initiation of S's action. Θ is entirely causally impotent with respect to S's ACE.

But if the neuroscientist were to ensure that Θ is entirely causally impotent vis-à-vis S's directly free action, then Θ would not contribute causally to S's ACE. Nevertheless, if Θ is inert as F&B stipulate, Θ cannot be a direct necessary causal condition of S's ACE. If Θ were inert as F&B stipulate, Θ would not imply an event that is a necessary causal condition of the ACE, in which case Θ would not imply an event that enables the ACE. But condition (5) requires that Θ contribute causally to S's ACE. In short, condition (5) is inconsistent with F&B's stipulations. Hence, strictly speaking, F&B's counterexample misfires, as they have not satisfied O'Connor's conditions in the first place.[20]

Intuitively, for finite creatures such as ourselves, purposive actions require some or other motivation. That is, purposive actions require some or other reasons in the form of beliefs and desires. If this is right, then for every ACE, the agent acts on some or other reason. And O'Connor tells a story about how it could be that a particular reason in some sense moves an agent to act freely. *If* (i) O'Connor's four conditions for reasons-explanation are satisfied; (ii) a certain reason contributes causally to the agent's self-determining activity (i.e., her ACE);[21] and (iii) this reason confers, prior to the agent-causal event's occurrence, an objective propensity for the agent to initiate her action by agent-causing an intention, *then* the agent freely acts *for* this reason—the reason *moves* the agent to act.

---

[18] Say that C is a *direct* necessary causal condition of E just if C is a necessary causal condition of E and there is no F such that C is a necessary causal condition of F and F is a necessary causal condition of E.

[19] For example, Θ is not itself a necessary causal condition of the ACE but Θ still implies a certain event D that is a necessary causal condition of the ACE. For example, D may be an essential part of the neural realization of Θ. Or, perhaps (i) D is a part of Θ, (ii) D is a necessary causal condition of some event E, where (iii) the ACE is an essential part of E, and (iv) D is not a necessary causal condition of any proper part of E. While these examples are illuminating, certain other restrictions may apply and are handled by my account.

[20] It is worth noting that even if one rejects my account of causal contribution, clearly something could contribute causally to an event even though nothing produces this effected event. There is a difference between production and mere-causal influence.

[21] More precisely, the agent's having the reason implies an event D that enables a minimal canopy condition E* for the performance of the action. See Chapter 6.

O'Connor's view makes even more sense when we reflect on the fact that we often have reasons that motivate and move us without our acting on those reasons. This is a trivial case of reasons inclining without necessitating. Suppose that I'm inclined, moved, and motivated to drink the cup of espresso at the end of dinner. Nonetheless, I freely decide to abstain, freely opting for a glass of milk instead. Prior to the choice of milk (over espresso), I have other inclinations and tendencies besides those that favor deciding on espresso. All of my inclinations and desires move me. That is just what properly functioning inclinations and desires do. Moreover, frequently I *feel* differences in the strength of my inclinations or desires, which push or pull me one way or another. Nonetheless, I might freely act for only some of those reasons.

Is it possible that the desire for espresso moves me to choose the milk? It may be possible, but neither is it clearly the case that I would then be acting freely nor is it clearly the case that O'Connor's conditions would then be satisfied. We may concede that the desire for espresso moves me in the weak sense that I feel the urge to pick espresso. Yet I select milk. But we are not entitled to infer that the desire for espresso moves me *to* select milk, rather than simply moving me *in* or *during* my selection of milk. Why? According to O'Connor's conditions, in virtue of what do I select milk *for* the reason that I desire milk and not *for* the reason that I desire espresso? Why privilege the desire for milk over the desire for espresso?

The answer is straightforward. My having the desire for milk contributes causally to my selecting milk—likely because my decision that begins my action of selecting milk nomologically requires my having the desire for milk (or an event included by my desire for milk). More precisely, it is likely that my desire for milk implies some event that enables a minimal canopy condition for my decision that begins my action of selecting milk. So my having the desire for espresso does not, but my having the desire for milk does, contribute causally to my decision. Moreover, prior to my decision, the desire for milk induces or elevates an objective propensity of my decision that begins my action of selecting milk. But the same does not hold for my having the desire for espresso. Prior to my decision, my having the desire for espresso induces or elevates an objective propensity of some other decision—presumably, a decision that would begin my selecting espresso instead. Yet even though there was in me some propensity to select espresso, I freely opted for milk.

8.6   Against Using My Account of Causal Contribution
I proposed an account of causal contribution in Chapter 6. My account, to review, is chiefly inspired by Roderick Chisholm's theory. In Chapter 6 I postponed a few objections claiming that Chisholm's theory cannot handle a reason's causally influencing an agent's free action. If Chisholm's theory fails, one might be tempted to think that mine fails for the same sort of reason. Moreover, since this chapter relies on the claim that an agent's reasons may contribute causally to her directly free action, it would be appropriate to consider whether my general account of causal contribution falls prey to counterexample.

Interestingly, even O'Connor raises an objection against Chisholm's analysis of causal contribution, and so one might be tempted to think that it threatens my account as well. I'll argue that the success of O'Connor's objection violates another one of his commitments. I'll also argue that O'Connor's objection is, despite appearances, not a problem for my account.

It is important to note that even if my account of causal contribution fails, assuming that there is such a thing as mere-causal-contribution is still open to the AC theorist. Even if we do not presently have an adequate theory of causal contribution, every fully adequate theory of free

will presumes that there is one. So, failing to have a satisfactory account is not a special problem for the AC proponent. The AC theorist, for example, may take as primitive the notion of causal contribution in his response to F&B's criticism—see 8.5.2. At the same time, providing an account of causal contribution and seeing that it confirms the AC theorist's fundamental commitments underscores the credibility of our attempt to render agent-causation intelligible.

O'Connor (2002b) agrees with Chisholm's assumption that, possibly, there is a free action having no sufficient causal condition. In such a case, an event contributes causally to the free action in virtue of the free action's having some necessary causal condition. O'Connor (2002b: 329) explains Chisholm's view thus:

> Chisholm suggests that reasons are necessary causal conditions on agent's causing their actions. I am always acting with some purpose, and my desiring to attain that end and having appropriate beliefs about how to do so thereby contribute to my doing so, not by forming part of a sufficient [causal] condition for the action, but by their being essential to *preventing* the occurrence of a sufficient condition for my *not* causing the action. (Had those factors not obtained, there would have been a sufficient [causal] condition for my not causing that particular action.)[22]

O'Connor (2002b: 349) constructs a putative counterexample:

> One problem with this way of understanding the role of reasons is that we can envision cases where my having reason [R] and my having reason [R*] each guide my [freely] performing an action but neither of which is such that, had that state not obtained, I would have been precluded from [freely] performing the action.

I'd like to make four points in response to O'Connor's criticism. First, the success of O'Connor's objection is inconsistent with his (1995b) claim that an agent's reasons are necessary causal conditions of her directly free action—i.e., of her ACE.[23] Indeed, in the very same essay including the quotation just cited, O'Connor (2002b: 352) speaks of "the very attenuated role O'Connor [1995b] assigns [to reasons], in which my now having a reason to A is a necessary causal condition on my now causing the intention to A." Thus, at least O'Connor once held that a reason causally influences a free action by being a NCC of the action. Moreover, O'Connor seems not to have changed his mind. Recall from Chapter 5 that O'Connor believes that while nothing can produce/trigger an ACE, the ACE may have a structuring cause. Since a structuring cause C of ACE ($\Delta\bullet\rightarrow e$) establishes a necessary context in which Δ exerts her characteristic purposive power in directly producing e, it seems quite reasonable to think that C is a NCC of the ACE.[24] So, at the very least, the AC theorist is not unreasonable to think of O'Connor's structuring condition of an ACE as implying a NCC of the ACE.

---

[22] I've inserted the term 'causal' since Chisholm is clearly not referring to broadly logically sufficient conditions for the action but rather to a sufficiency of a nomic modality.

[23] Hereafter, let "NCC" and "SCC" denote "necessary causal condition" and "sufficient causal condition," respectively.

[24] Quite recently O'Connor (personal correspondence, September 2004) explains:

Second, concerning O'Connor's thought-experiment, it may be that reasons R and R* both complete (or each implies an event that completes) a SCC C of a NCC N of the agent's ACE. This is consistent with O'Connor's stipulations that the reasons guide the agent in performing the action even though neither R nor R* is a NCC of the ACE. It may be that for each reason, if it had not occurred, there would have been some other SCC C* of the same NCC N of the agent's ACE. If this is right, it confirms both Chisholm's account and my account, for it describes another way in which two reasons guide performing an action while contributing causally to that action.

Admittedly, my response depends on the controversial assumption that the agent's reason contributing to her ACE via some intermediate event. That is, I assume that reasons R and R* do not *directly* contribute causally to the ACE. Suppose, though, that O'Connor simply stipulates this assumption away. That is, suppose we also stipulate that each reason directly guides the agent's free behavior. So, if each reason contributes causally to the agent's ACE, then the contribution is direct. This is probably the case O'Connor had in mind anyway. Nonetheless, once we stipulate that any causal contribution of the agent's reasons on her ACE is direct, the case is to some degree more difficult to imagine and proportionally undercuts its status as a clear counterexample. It may be that, necessarily, whichever laws of nature permit agent-causal events, an agent's reason contributes causally to her ACE only if her reason contributes to an event E that contributes to the ACE. Moreover, it may be that no such intermediate event E can be a reason, even though a reason may contribute causally to E. Fortunately, we need not make the controversial assumption that a reason guides and contributes causally to an action only if the contribution is indirect. This brings us to our third point.

Third, even though neither reason R nor reason R* is itself a NCC of the ACE, Chisholm's (1986) and my view of causal contribution only commits us to the following claim. Each reason directly contributes causally to an ACE only if each implies an event that is a NCC of the ACE. Since each of R and R* guides the agent in freely performing some action, it is reasonable to think that each implies the same event that plays the right causal role in the action. Thus, it is open to the AC theorist to think that R implies an event D that is a NCC of the free action while R* also implies D. In this case, the contribution of R and R* on the agent's free action is simultaneous, which confirms the claim that both guide the agent in freely performing the action.

Can we imagine a case where (i) reason R guides some finite agent Δ's performing some directly free action A, (ii) reason R* likewise guides Δ's performing A, (iii) neither R nor R* is a necessary causal condition of A, (iv) neither R nor R* implies an event D that is a necessary causal condition of A? At this point, it is not clear that we're imagining something possible.

> Now, I was also using Dretske's term ["structuring causes"] (without being explicit about this) in a somewhat wider sense, to refer to the causes of any standing conditions that are necessary for the operation of a given cause in the first place. So, the proper functioning of vital bodily systems, the engagement of my attention, the absence of severe, distracting pain, the presence of sufficient oxygen in my environment, and countless other factors all are, in this wider sense, structuring causes of any agent causal event. They are necessary conditions, but not jointly sufficient, and they do not trigger my causing an intention. Instead, they establish a context in which I am able to cause such an intention.

Clearly, the modality denoted by 'necessary' should not be that of broadly logical necessity but rather that of nomological/physical necessity.

The only positive qualification of the case is put in terms of a reason's guiding an agent's freely performing A. But what is this notion of guidance? More precisely, what is the three-placed relation *reason x guides S's performing action y*? Is it equivalent to the relation *S performs action y for reason x*? I'm inclined to think not.

For example, you are on a podium. I set the podium on fire. You are able to jump toward the north, east, south, or west. You notice that the terrain to the west is unfavorable. So you have a reason not to jump toward the west. You notice that all of the other terrain is favorable. You desire to jump toward the east, and you desire to jump toward the north. Finally, you freely jump toward the east for the reason that you wanted to jump toward the east. Although your desire not to jump toward the west guided your action, you did not act for that reason.

Thus, instantiating the relation *reason x guides S's performing action y* is not sufficient for instantiating the relation *S performs action y for reason x*. So, until we know more about the former relation, we should not be surprised to discover that it does not imply the relation *x's contributing causally to S's performing y*. Since the putative point of contention concerns whether the relation *S performs y for reason x* implies *x's contributing causally to S's performing y*, any clear counterexample should describe a case of an agent's acting for a reason and not a case of a reason's guiding the agent's acting.

My fourth and final point concerns O'Connor's characterization of Chisholm's view. To review, O'Connor (2002b: 349) claims, "Chisholm suggests that reasons are necessary causal conditions on agent's causing their actions." But to my knowledge Chisholm never claims that an agent freely acts and acts directly for a reason only if this reason is (or implies an event that is) a necessary causal condition of the free act. More generally, to my knowledge Chisholm never claims that an agent freely acts for a reason only if this reason contributes causally to the free act. Rather, Chisholm claims that, possibly, a reason contributes causally to a free action. When Chisholm speaks of reasons as contributing causes, Chisholm is in the process of answering an objection claiming that reasons cannot causally contribute to one's free action. So, he imagines a case where they do contribute causally to a free action. But there's no reason to think that his case represents the only way an agent can act for a reason.[25] Chisholm, then, should only be committed to thinking that a reason's contributing causally to an agent's free action can be part of informative *sufficient* conditions for the agent's freely acting for that reason. The same point applies to how I have used my account of causal contribution, as I too have attempted to provide informative, agent-causal sufficient conditions for an agent's acting freely for a reason. I conclude that there is no objection against using my account of causal contribution in elucidating how it can be that a prior event causally influences an agent's directly free action. In my view, the AC theorist has resources to render intelligible how an agent could act freely for a reason.

8.7   Conclusion on Freely Acting for Reasons
I have argued on behalf of O'Connor's sufficient conditions for a rational free action. After reviewing his account, I argued that Feldman & Buckareff's criticism failed to undermine

---

[25] My claim here confirms my first reply to F&B's objection against O'Connor's basic account. My first reply defended O'Connor's basic account from their putative counterexample without relying on the claim that the reason on which an agent freely acts contributes causally to the action.

O'Connor's explicit account. Moreover, I suggested that the agent-causal theorist has resources to respect the causalist's so-called intuition that reasons can causally influence one's directly free action. I noted textual evidence from O'Connor's work to motivate a modified account that explicitly mentions a reason's merely-contributing causally to an agent's directly free action. Finally, I suggested how one could wed my account of causal contribution with O'Connor's conditions, thereby illustrating what this modified account might look like. I conclude that agent-causal theorists can provide sufficient and informative conditions for an agent's acting freely for a reason. Objections claiming otherwise look dim. Indeed, I think that an AC theorist can make fairly decent sense of an undetermined and rational free action. Her theory seems quite plausible. Moreover, given Chapter 5's arguments for thinking that agent-causation best accounts for how an agent is an underived originator of her free behavior, I conclude that agent-causation is the best theory of free will available.

CONCLUSION

The prospects for agent-causation as a theory of directly free action look quite bright. Our journey to this conclusion consisted of eight chapters.

Chapter 1 outlined the significance of the metaphysical problem of free will, which roughly goes as follows. If an act is determined, then it is not free. If an act is not determined, then it is not free. But either an act is determined or not. So, a free act is impossible. However, it is most obvious that we could act freely. For example, it is clearly possible that we are morally responsible for things. Moreover, clearly it must be the case that one is morally responsible for something only if one could act freely. It follows that we could act freely. Chapter 1 framed our investigation by asking the following three questions. What kind of freedom is most worth wanting? Is this kind of freedom consistent with determinism? If not, how is it that this freedom is consistent with indeterminism?

Chapter 2 got very clear about our pre-theoretic concept of acting freely. For, we cannot begin to assess the merits of a solution to the metaphysical problem of free will unless we have some bedrock convictions about free action. Consulting how experts of free will characterize its pre-theoretic status suggests that there is a strict and demanding kind of free will that at least requires the following two conditions. First, one acts freely only if one could have refrained freely (perhaps by freely doing otherwise)—I called this the *condition of dual power*. Second, one acts freely only if one is an underived source or originator of one's behavior—I called this the *origination condition*.[26]

Chapter 3 discussed in detail the hackneyed dispute about whether or not free will is compatible with determinism. I got clear about the meaning of 'determinism'. I suggested that ordinary folk already believe that acting freely precludes determinism. So, I proposed that we may reasonably believe without further argument that they are incompatible until we encounter a good argument to think otherwise. Nonetheless, I investigated the merits of Ted Warfield's (2000) argument for incompatibilism. Though I defended his argument from Dana Nelkin & Samuel Rickless's (2002) objection, I sided with Tom Flint (1998b), arguing that Warfield's argument fails for other reasons. Chapter 3 concluded with the modest claim that even if our default conviction in incompatibilism is mistaken, it may also be that an undetermined free act is possible as well—after all, compatibilism is only a compatibility thesis. Thus compatibilists and incompatibilists alike should take seriously the task of rendering intelligible an undetermined free act.

But maybe there cannot be an undetermined free action. Chapter 4 advanced a reasoned and thorough critique of Peter van Inwagen's (2001, 2000) argument for thinking that an undetermined free act is impossible. If I'm right, there is one less barrier to taking seriously prospective theories of undetermined free action.

Chapter 5 was a pivotal chapter. I evaluated the three main candidates for understanding the nature of an undetermined free act: Carl Ginet (1997, 1990) and Stewart Goetz's (1997)

---

[26] Given the condition of origination, one may adjust the intuitive condition of dual power as follows. One acts freely only if one could have refrained from being an originator of an event (perhaps by being an originator of a different event instead).

simple indeterminism, Robert Kane's (1996a) causal indeterminism, and Timothy O'Connor's (2000a) agent-causation. I argued in detail that of these three views, only agent-causation clearly meets the condition of origination (*viz.*, Whenceness) that free will requires. The signature of agent-causal theories is that an agent *qua* enduring entity strictly and literally makes occur (i.e., produces) an essential element of her free action. I drew an intuitive distinction between full-fledged causal production, on the one hand, and mere-causal-contribution, on the other hand. Chapter 5 presented several arguments for thinking that, given an agent-causalist understanding of what is an agent's directly free action, nothing can causally produce an agent's directly free action. I contended that no other theory of free action gives a principled account for thinking that nothing could produce an agent's directly free action, which is a consequence that nicely confirms our pre-theoretic commitment to incompatibilism. Chapter 5 concluded that since agent-causation uniquely meets the origination condition, there is a defeasibly good reason to prefer agent-causation over both simple indeterminism and causal indeterminism. The reason is defeasible because the vices of agent-causation may outweigh its virtue in meeting the origination condition.

Since I argued in Chapter 5 that nothing can produce an agent's directly free action even though something might merely-contribute causally to her directly free action, it would be nice to have an account of how an event contributes causally to another event. Nice, but not necessary. Nonetheless, chapter 6 advanced such an account of causal contribution. My account of causal contribution is similar to Roderick Chisholm's (1986), but mine overcomes a host of difficulties that plague his theory.

Chapter 7 entertained about ten of the most frequently cited criticisms of the agent-causal view. Once charitably cast in the form of an argument, we saw that none of these criticisms threatens agent-causation.

Chapter 8 considered the question *Can the agent-causalist give sufficient and informative conditions for an agent's acting freely for reasons*? O'Connor (2000a) thinks so, and I explained his account. But Richard Feldman & Andrei Buckareff (2003) contend that O'Connor fails. I criticized their contention. Moreover, I showed that even if their counterexample sticks, one can repair O'Connor's conditions by appealing to the notion of mere-causal-contribution.

I conclude that agent-causation is the best theory for making sense of an undetermined free act, since it best meets the conditions that free will clearly requires and objections against it look rather dim.

APPENDIX 1

CAUSAL CONTRIBUTION

Chisholm's (1986: 59-64) Account—see figure 1:

(1)   Event C is a *sufficient causal condition* of event E =df. It is not logically necessary that if C occurs, then E occurs; and it is physically necessary that if C occurs, then E occurs either at that same time or later.

(2)   Event C is a *necessary causal condition* of event E =df. It is not logically necessary that if E occurs then C occurs; and it is physically necessary that if E occurs, then C occurs either at that same time or earlier.

(3)   Property P is *conjunctive* =df. There are two further properties which are such that (a) neither implies the other, (b) P implies each of them, and (c) P is implied by whatever implies both of them.

(4)   Event C *completes* a sufficient causal condition of event E =df. There occurs a set of events such that (a) none implies a conjunctive property and (b) their conjunction is a sufficient causal condition of E; and C is a member of every such set.

(5)   Event C *contributes causally* to event E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E or (b) E occurs and D completes a sufficient causal condition of a necessary causal condition of E.
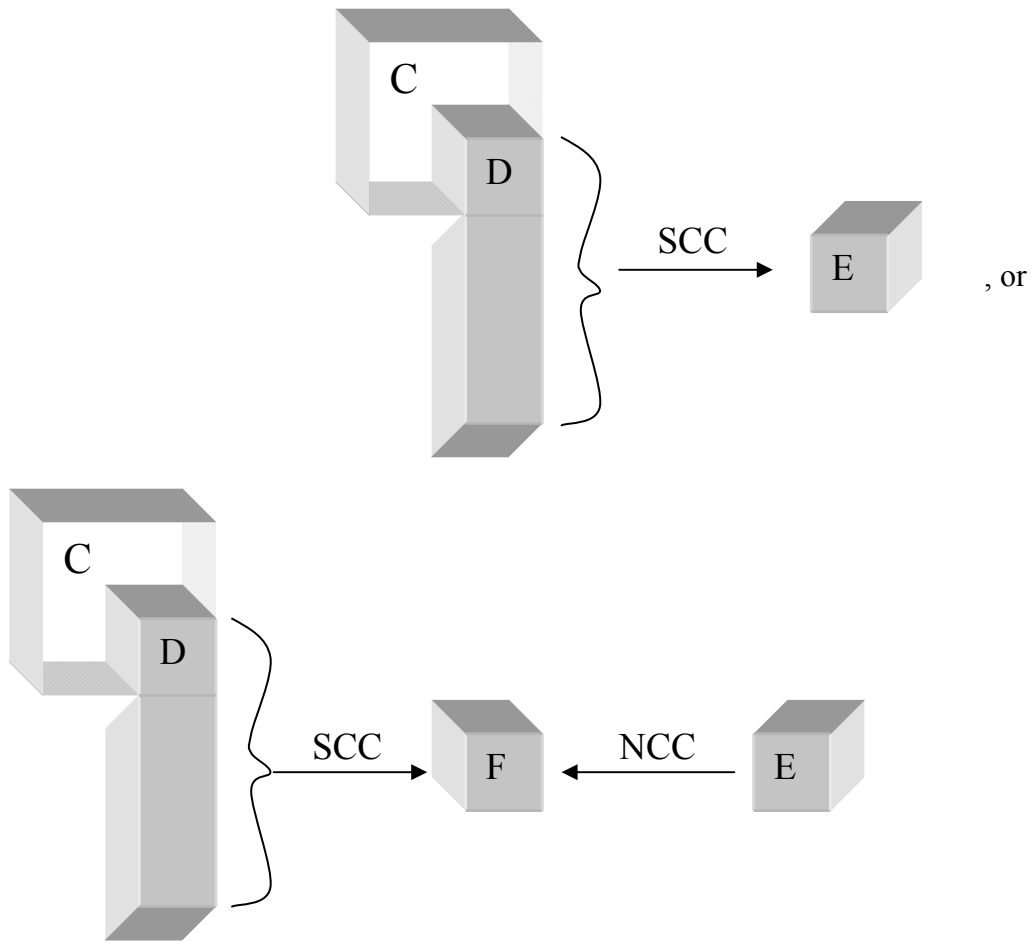
Figure 1: Chisholm's (1986) Account

My account is as follows (see figure 2):

(1)   Event C is a *sufficient causal condition* of event E =df. It is not logically necessary that if C occurs, then E occurs; and it is physically necessary that if C occurs, then E occurs either at that same time or later.

(2)   Event C is a *necessary causal condition* of event E =df. It is not logically necessary that if E occurs then C occurs; and it is physically necessary that if E occurs, then C occurs either at that same time or earlier.

(3)   Property P is *conjunctive* =df. There are two further properties which are such that (a) neither implies the other, (b) P implies each of them, and (c) P is implied by whatever implies both of them.

(4*)   Event C *completes* a sufficient causal condition of event E =df. There occurs a set S of events such that (a) none implies a conjunctive property, (b) their conjunction is a sufficient causal condition of E, and (c) no conjunction of any proper subset S* of S is a sufficient causal condition of E; and C is a member of some such set S.

(6)   Event D *enables* E *by a minimal canopy condition* E* =df. E occurs; and there is a set S of events such that (i) none of S's members implies a conjunctive property and the conjunction of the members of S is equivalent to event E*; (ii) E* implies and is contemporaneous with E; (iii) D either is a necessary causal condition or completes a sufficient causal condition of a necessary causal condition of E*; and (iv) D neither is a necessary causal condition nor completes a sufficient causal condition of a necessary causal condition of a proper subset of S.

(7)   C is a *global causal condition* of E =df. C occurs; and C implies an event D such that either (a) D completes a sufficient causal condition of E, or (b) D enables E by a minimal canopy condition E*. (See figure 2.)

(8)   C is an *atomic* global causal condition =df. C is a global causal condition implying no conjunctive property.

(9)   Event C *contributes causally* to event E =df. C is a global causal condition of E; and every atomic global causal condition of C is a global causal condition of E.

213

C

D

SCC → E , or

E* = atomic events

C

NCC

E*

D

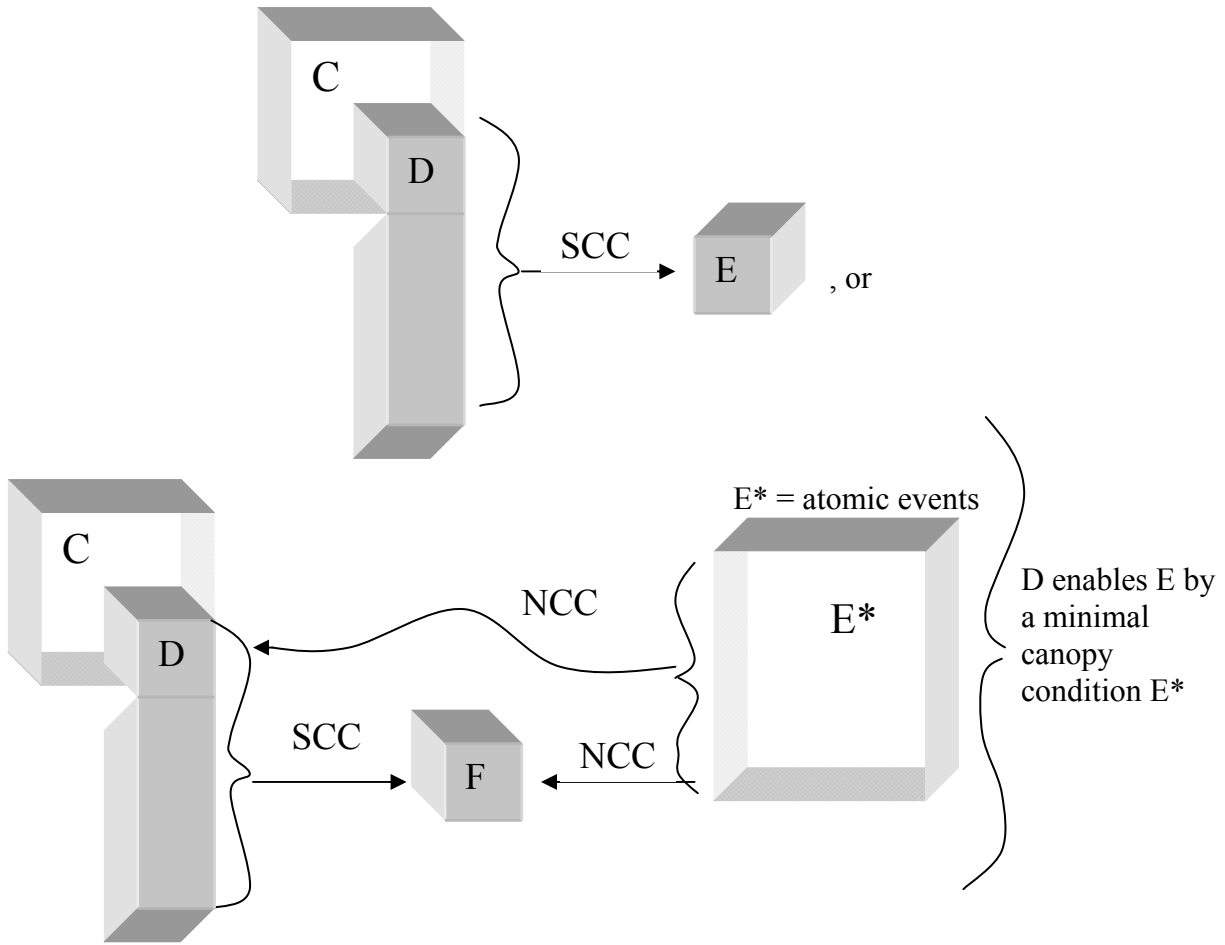D enables E by a minimal canopy condition E*

SCC

F

NCC

Figure 2: C is a *global causal condition* of E

A bit more chisholming? Suppose we remove a disjunct from the enabling a minimal canopy condition. We then need to build transitivity into our final account. To do this, add (10) and replace (6) and (9) with (6*) and (9*), respectively—see figure 3:

(6*)   Event D *enables* E *by a minimal canopy condition* E* =df. E occurs; and there is a set S of events such that (i) none of S's members implies a conjunctive property and the conjunction of the members of S is equivalent to event E*; (ii) E* implies and is contemporaneous with E; (iii) D is a necessary causal condition of E*; and (iv) D is not a necessary causal condition of a proper subset of S.

(10)   Event C *directly* contributes causally to event E =df. C is a global causal condition of E; and every atomic global causal condition of C is a global causal condition of E.

(9*)   Event C *contributes causally* to event E =df. C is a member of some *thread* to E, where a thread to E is an ordered n-tuple $<C_n,….,C_1>$ for any $n \geq 1$ such that both (i) C_1 directly contributes causally to E and (ii) for every pair of members $C_{(n+1)}$ and C_n, $C_{(n+1)}$ directly contributes causally to C_n.
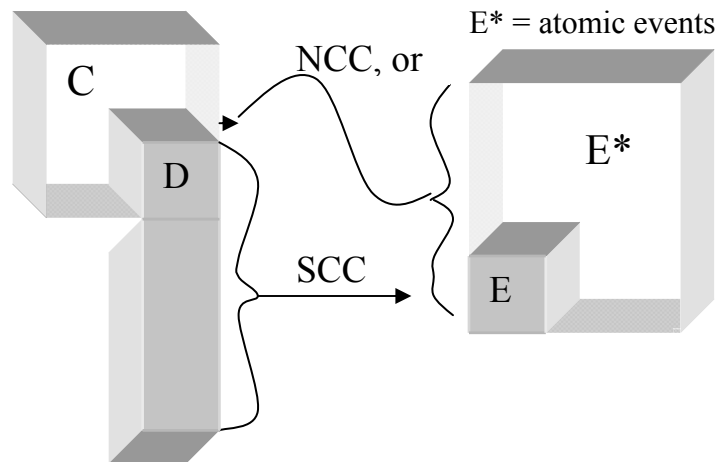

Figure 3: Finished Account

APPENDIX 2

O'CONNOR'S TEXTS ON NOTHING'S PRODUCING AN ACE

Here are O'Connor's reasons for thinking that it is impossible to produce an agent-causal event (ACE). O'Connor (1995b: 186) reasons:

> The cause of *A's causing of B* is none other than the cause of A itself. [O'Connor inserts this note: "Assuming, that is, that what we are after is the 'triggering' cause of the event, rather than what Fred Dretske [1988] calls a 'structuring' cause—roughly, that which establishes a causal pathway between two objects or systems so that when the first is operated upon (by the triggering cause) in the right manner, it brings about a result in the latter."] What, then, of *S's causation of e*? There appears to be no way of getting a grip on the notion of an event of *this* sort's having a sufficient, efficient cause. Because of its peculiar causal structure, there is no event at its front end, so to speak, but only an enduring agent. And there cannot be an immediate, efficient cause of a causal relation (i.e., independently of the causation of its front end relatum). In general, that which is causally produced in the first [i.e., a metaphysically basic] instance is always an event or state having a causally *simple* structure: an object O's exemplifying *intrinsic* properties p_1, p_2, … at time t_0. Causally complex events can also be caused, of course, but only in a derivative way: where they have the form *event X's causing event Y*, whatever causes event X is a cause *thereby* of *X's causing Y*. In the special case of an *agent's* causing an event internal to his action, however, there is no causally simple component event forming its initial segment, such that one might cause the complex event (*S's causing e*) in virtue of causing *it*. Therefore, it is problematic to suppose that there could *be* [an efficient cause that makes occur] an agent-causal event.

O'Connor (1996: 147) argues:

> For instantiations of causal *relations* (causally complex events) are not themselves directly on the receiving end of other causal relations—instead, instantiations of intrinsic properties (causally simple *states* or events) are. Causings are the *producing* of events, rather than what are *produced* (in the first [i.e., metaphysically basic] instance). Compare an ordinary case of an event-causal process (consisting of event *F*'s causing event *G*) being caused by some further *event E*. Surely this can consist only in *E*'s causing *F*, the front-end relatum of the complex event. [O'Connor inserts a note advocating Dretske's (1988: Chapter 2) distinction between structuring causes and triggering causes.] When I reflect on the matter, I cannot but regard this as reflecting an evident, *general* truth about

causation. If this is right, then an *agent*-causal event could not be caused for the simple reason that the cause in this case is not itself an event.

O'Connor (2000a: 53) reasons,

> Consider a familiar sequence of events. My finger presses a doorbell button, the doorbell rings, and your cat jumps in fright. We may sensibly say that my finger's pressing the button causes the causal sequence, *the ringing of the bell's causing the cat to jump*. But what we mean here is simply that it caused the sequence indirectly, by causing the first element of the sequence, the bell's ringing. We may also sensibly say that the electrician's wiring of the doorbell system was a cause of the sequence, *the depressing of the button's causing the bell to ring*. Following Fred Dretske [1988: Chapter 2], we may term this latter episode one of 'structural' causation, which consists in establishing a causal pathway—here, the wiring and power supply—between two objects or systems that is subsequently triggered by some appropriate event. Here we mean only that the 'structuring' cause provided a context in which some causal factor exerted its characteristic effect. It is not to say that the establishment of an electrical pathway in any way brought about or enabled the button's depression's exerting its characteristic influence on its immediate environment, only that it will determine one important wider effect of that influence.
> Neither of these legitimate ways of speaking of causes of causings within event-causal contexts supports the idea of a cause of an agent's causing his own intention. The first type of example has no analogue in the context of agent causation because the cause within the causally complex event, *agent S's causing e*, is not itself an event but an enduring substance. The second example, involving structuring causes, is clearly applicable to free human agency. (The wider consequences of our immediate effects is structured by countless external factors) Yet it is irrelevant to the idea that some independent factor might directly bring about the causal activity of any basic cause, whether of the agent or event variety. Hence, we should reject this idea altogether.

O'Connor (2000a: 61) reasons that

> [R]eflection on the nature of [an ACE] suggest that there cannot, in the nature of the case, be a cause that produces it. (Undoubtedly, a wide range of factors is [causally] necessary for an agent to have such a causal capacity at the time of its exercise.) First, that something should bring about [i.e., directly produce] the causing of an event is absurd: it implies that the cause in the basic causal transaction is in some way deficient, that something further is needed to bring about its causing of its effect. Second, there is no principled way to stop at just one level of metacauses. Positing an infinity of such causes of causings is not merely ridiculous on the face of it but also logically vicious.

O'Connor (2002a: 135-6) reasons:

> I argue that an ACE cannot have a triggering-cause—an event that stimulates the cause (in this case, me, a substance) into action. Consider first event causes. Where agent causes are absent, they are linked in constant flow of one event's giving rise to another. Now, where event B causes C, there is not a direct triggering cause of *B's causing C*. Instead, some event A brings about B, which brings about C, and so on. (We might say, if we like, that A indirectly triggers *B's causing C* in virtue of triggering A.) In the case of an agent-causal event, however, there is no front-end event to be caused, only an agent qua substance. So there doesn't seem to be room, as a simple conceptual matter, for an ACE to have a triggering cause.

REFERENCES

Allen, Robert (1997) 'Re-examining Frankfurt Cases' *Southern Journal of Philosophy* 37: 363-376.

Anscombe, G.E.M. (1998) 'Causality and Determination' in van Inwagen and Zimmerman, eds. (1998: 244-257). Originally printed in (1971) *Inaugural Lecture* (Cambridge: Cambridge University Press). Also reprinted in Anscombe (1981: 57-73).

Anscombe, G.E.M., ed. (1981) *The Collected Papers of G.E.M. Anscombe*, volume 2 (Minneapolis: University of Minneapolis press).

Aquinas, Thomas (1945) *Basic Writings of Saint Thomas Aquinas*, volume I, Anton Pegis, ed. (New York: Random House).

Aristotle (1984) *The Complete Works of Aristotle: the Revised Oxford Translation*. Jonathan Barnes, trans. (Princeton, NJ: Princeton University Press).

Armstrong, David (1997) *A World of States of Affairs* (New York: Cambridge University Press).

Armstrong, David (1983) *What Is a Law of Nature?* (Cambridge: Cambridge University Press).

Armstrong, David (1978) *Universals and Scientific Realism* (Cambridge: Cambridge University Press).

Augustine (1964) *On the Free Choice of the Will* (Indianapolis: Bobbs-Merrill).

Austin, Jane (1961) 'Ifs and Cans' in his *Philosophical Papers* (Oxford: Clarendon Press).

Ayer, A.J. and J. O'Grady, eds. (1992) *A Dictionary of Philosophical Quotations* (Oxford: Blackwell Publishers).

Ayer, A.J. (1954) 'Freedom and Necessity' *Philosophical Essays* (London: Macmillan).

Ayer, M.R. (1968) *The Refutation of Determinism* (London: Methuen).

Balashov, Yuri (2002) 'What is a Law of Nature? The Broken-Symmetry Story' *Southern Journal of Philosophy* 40: 459-73.

Barker, Stephen and Tom Beauchamp, eds. (1976) *Thomas Reid: Critical Interpretations* (Philadelphia: University City Science Center).

Basinger, David (1987) 'Middle Knowledge and Human Freedom: Some Clarifications' *Faith and Philosophy* 4: 330-336.

Beaty, Michael, ed. (1990) *Christian Theism and the Problems of Philosophy* (Notre Dame, IN: University of Notre Dame Press).

Benson, Paul (1987) 'Freedom and Value' *Journal of Philosophy* 84: 465-486.

Berofsky, Bernard (1987) *Freedom from Necessity* (London: Routledge & Kegan Paul).

Berofsky, Bernard, ed. (1966) *Free Will and Determinism* (New York: Harper and Row).

Bertolet, Rod (1993) 'Hasker on Middle Knowledge' *Faith and Philosophy* 10: 3-17.

Beyer, Richard (2003) *The Greatest Stories Never Told* (New York: HarperCollins Publishers).

Bigelow, John, Brian Ellis, and Caroline Lierse (1992) 'The World as One of a Kind: Natural Necessity and Laws of Nature' *British Journal for the Philosophy of Science* 43: 371-88.

Binkley, Robert (1965) 'A Theory of Practical Reason' *Philosophical Review* 76: 423-448.

Bishop, John (1989) *Natural Agency: An Essay on the Causal Theory of Action* (Cambridge: Cambridge University Press).

Bishop, John (1986) 'Is Agent-Causality a Conceptual Primitive?' *Synthese* 67: 225-247.

Bradie, Michael and Myles Brand, eds. (1980) *Bowling Green Studies in Applied Philosophy*, volume 2 (Bowling Green, OH: Bowling Green State University).

Bramhall, John (1844) *The Works of John Bramhall*, volume 4 (Oxford: John Henry Parker).

Brand, Myles (1984) *Intending and Acting* (Cambridge, MA: MIT Press).

Brand, Myles and David Walton, eds. (1976) *Action Theory* (Dordrecht-Holland: D. Reidel Publishing Company).

Broad, C.D. (1952) 'Determinism, Indeterminism and Libertarianism' in his (1952)

Ethics and the History of Philosophy (London: Routledge & Kegan Paul).

Brody, Baruch, ed. (1969) 'Introduction' *Essays on the Active Powers of the Human Mind*, by Thomas Reid (Cambridge, MA: MIT Press): vii-xxiv

Brown, Harold (1977) *Perception, Theory, and Commitment* (Chicago: University of Chicago Press).

Campbell, C.A. (1967) *In Defense of Free Will* (London: George Allen and Unwin, Ltd.).

Campbell, C.A. (1957) *On Selfhood and Godhood* (London: George Allen and Unwin, Ltd.).

Campbell, C.A. (1966) 'Is "Free Will" a Pseudo-Problem?' in Berofsky, ed. (1966: 112-135). First printed in (1951) *Mind* 60: 446-465.

Chisholm, Roderick (1995) 'Agents, Causes, and Events' in O'Connor (1995: 95-100).

Chisholm, Roderick (1986) 'Self-Profile,' in Radu J. Bogdan, ed. (1986: 59-64) *Roderick Chisholm* (Dordrecht: D. Reidel Publishing Company).

Chisholm, Roderick (1982) 'Human Freedom and the Self' in Watson (1982: 24-35) and Chisholm (1976). First published (1964) The Lindley Lecture (University of Kansas: Department of Philosophy): 3-15.

Chisholm, Roderick (1981a) 'Lawlike and Non-lawlike Generalizations' in *Philosophie als Wissenschaft/Essays in Scientific Philosophy*, Edgar Morscher, et al., eds. (Germany: Bad Reichenhall): 57-64.

Chisholm, Roderick (1981b) 'Laws of Nature and Causation' *Philosophische Forschung*: 57-61.

Chisholm, Roderick (1979) 'Objects and Persons: Revisions and Replies,' in Sosa, ed., (1979: 317-88) *Essays on the Philosophy of Roderick Chisholm* (Amsterdam: Rodopi).

Chisholm, Roderick (1978) 'Comments and Replies' *Philosophia* 7: 597-634.

Chisholm, Roderick (1976a) *Person and Object* (Lasalle, IL: Open Court).

Chisholm, Roderick (1976b) 'The Agent as Cause' in Myles Brand and Douglas Walton, eds. (1976: 199-211).

Chisholm, Roderick (1971) 'Reflections on Human Agency,' *Idealistic Studies* 1: 33-46.

Chisholm, Roderick (1967) 'He Could Have Done Otherwise' *Journal of Philosophy* 6: 409-417.

Chisholm, Roderick (1966) 'Freedom and Action' in Keith Lehrer, ed. (1966: 11-44).

Chisholm, Roderick (1958) 'Responsibility and Avoidability' in Sidney Hook, ed. *Determinism and Freedom in the Age of Modern Science* (New York: Collier-Macmillan).

Clarke, Randolph (2003) *Libertarian Accounts of Free Will* (New York: Oxford University Press).

Clarke, Randolph (2002) 'Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency' in Kane (2002a: 356-385).

Clarke, Randolph (2000) 'Modest Libertarianism' *Philosophical Perspectives* 14: 21-45.

Clarke, Randolph (1999) 'Free Choice, Effort and Wanting More' *Philosophical Explorations* 2: 20-41.

Clarke, Randolph (1997) 'On the Possibility of Rational Free Action' *Philosophical Studies* 88: 37-57.

Clarke, Randolph (1996a) 'Agent Causation and Event Causation in the Production of Free Action' *Philosophical Topics* 24: 19-48.

Clarke, Randolph (1996b) 'Contrastive Rational Explanation of Free Choice' *The Philosophical Quarterly* 46: 185-201.

Clarke, Randolph (1995) 'Indeterminism and Control' *American Philosophical Quarterly* 32: 125-138.

Clarke, Randolph (1993) 'Toward a Credible Agent-Causal Account of Free Will' *Nous* 27: 191-203. Reprinted in O'Connor (1995: 201-215).

Clarke, Samuel (1738) *The Works* (New York: Garland Publishing)

Copp, David (1997) 'Defending the Principle of Alternative Possibilities: Blameworthiness and Moral Responsibility' *Nous* 31: 441-456.

Craig, William (1998) 'On Hasker's Defense of Anti-Molinism' *Faith and Philosophy* 15: 236-240.

Craig, William (1994) 'Robert Adams's New Anti-Molinist Argument' *Philosophy and*

Phenomenological Research 54: 857-861.

Craig, William (1991) *Divine Foreknowledge and Human Freedom*. Brill's Studies in
    Intellectual History 19 (Leiden: E.J. Brill).

Craig, William and J.P. Moreland, eds. (2000) *Naturalism: A Critical Analysis*
    (London: Routlege).

Crisp, Thomas and Ted Warfield (2000) 'The Irrelevance of Indeterministic
    Counterexamples to Principle Beta' *Philosophy and Phenomenological Research* 61:
    173-84.

Davidson, Donald (1980) 'Agency' in his *Essays on Actions and Events* (Oxford:
    Oxford University Press).

Davidson, Donald (1973) 'Freedom to Act' in Ted Honderich, ed. (1973: 137-56).

Davidson, Donald (1963) 'Actions, Reasons, and Causes' reprinted in Davidson (1980)
    Essays on Actions and Events (Oxford: Oxford University Press): 3-20.

Dekker, Eef (2000) *Middle Knowledge* (Leuven: Peeters).

Dennett, Daniel (1995) 'On Giving Libertarians What They Say They Want' in
    Timothy O'Connor (1995a: 43-55). First printed in Dennett (1978).

Dennett, Daniel (1978) *Brainstorms* (Cambridge, MA: Bradford Books).

Donagan, Alan (1989) *Choice: The Essential Element in Human Action* (New York:
    Routledge & Kegan Paul).

Donagan, Alan (1979) 'Chisholm's Theory of Agency' *Grazer Philosophische Studien*:
    215-229.

Double, Robert (1991) *The Non-Reality of Free Will* (New York: Oxford University
    Press).

Dretske, Fred (1988) *Explaining Behavior: Reasons in a World of Causes* (Cambridge:
    MIT Press).

Dretske, Fred (1977) 'Laws of Nature' *Philosophy of Science* 44: 248-268.

Earman, John (1986) *A Primer on Determinism* (Dordrecht: Reidel).

Ekstrom, Laura (2003) 'Free Will, Chance, and Mystery' *Philosophical Studies* 113:
    153-180.

Ekstrom, Laura, ed. (2001a) *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, CO: Westview Press).

Ekstrom, Laura (2001b) 'Indeterminist Free Action' in Ekstrom, ed. (2001a).

Ekstrom, Laura (2000) *Free Will: A Philosophical Study* (Boulder, CO: Westview).

Ekstrom, Laura (1998) 'Freedom, Causation and the Consequence Argument' *Synthese* 115: 333-354.

Ekstrom, Laura (1995) 'Causes and Nested Conditionals' *Australasian Journal of Philosophy* 73: 574-578.

Ellis, Brian (2001) *Scientific Essentialism* (Cambridge: Cambridge University Press).

Ellis, Brian and Caroline Lierse (1994) 'Dispositional Essentialism' *Australasian Journal of Philosophy* 72: 27-45.

Fales, Evan (1990) *Causation and Universals* (London: Routledge).

Farren, Mick (2004) *Words of Wisdom* (Malaysia: Chrysalis).

Farrer, Austin (1967) *The Freedom of the Will* (New York: Scribner's).

Feldman, Richard and Andrei Buckareff (2003) 'Reasons Explanations and Pure Agency' *Philosophical Studies* 112: 135-45.

Feinberg, Joel and Russ Shafer-Landau, eds. (2002) *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, 11th edition (Stamford, CT: Wadsworth).

Finch, Alicia (2002) *Freedom and Incompatibilism: On the Possibility of Undetermined Free Action*, Ph.D. Dissertation, University of Notre Dame.

Finch, Alicia and Ted Warfield (1998) 'The *Mind* Argument and Libertarianism' *Mind* 107: 515-28.

Fischer, John (2000) '*The Significance of Free Will* by Robert Kane' *Philosophy and Phenomenological Research* 40: 141-8.

Fischer, John (1999) 'Recent Work in Moral Responsibility' *Ethics* 110: 93-139.

Fischer, John (1995) 'Libertarianism and Avoidability: A Reply to Widerker' *Faith and Philosophy* 12: 119-125.

Fischer, John (1994) *The Metaphysics of Free Will: an Essay on Control* (Oxford:

Blackwell Publishers).

Fischer, John, ed. (1986a) *Moral Responsibility* (Ithaca: Cornell University Press).

Fischer, John (1986b) 'Introduction: Responsibility and Freedom' in Fischer (1986: 9-61).

Fischer, John (1986c) 'Power Necessity' *Philosophical Topics* 14: 77-91.

Fischer, John (1982) 'Responsibility and Control' *Journal of Philosophy* 79: 24-40.

Fischer, John and Mark Ravizza (1998) *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press).

Flint, Thomas (2005) 'Regresses and the Theory of Agency,' unpublished manuscript.

Flint, Thomas (2001) 'The Possibilities of Incarnation: Some Radical Molinist Suggestions' *Religious Studies* 37: 307-320.

Flint, Thomas (1998a) *Divine Providence: the Molinist Account* (Ithaca: Cornell University Press).

Flint, Thomas (1998b) 'Comments on Fritz Warfield's "Causal Determinism and Human Freedom are Incompatible: Two New Arguments for Incompatibilism,"' personal manuscript presented at the University of Notre Dame's Philosophy Colloquium.

Flint, Thomas (1991) 'Hasker's *God, Time, and Knowledge*' *Philosophical Studies* 60: 103-115.

Flint, Thomas (1988) 'Two Accounts of Providence' in Thomas Morris, ed., *Divine and Human Action* (Ithaca: Cornell University Press): 147-181.

Flint, Thomas (1987) 'Compatibilism and the Argument from Unavoidability' *Journal of Philosophy* 84: 423-440.

Foley, Richard (1979) 'Compatibilism and Control Over the Past' *Analysis* 39: 70-74.

Frankfurt, Harry (1988) *The Importance of What We Care About* (Cambridge: Cambridge University Press).

Frankfurt, Harry (1969) 'Alternate Possibilities and Moral Responsibility' *Journal of Philosophy* 66: 829-839. Reprinted in Fischer (1986a: 143-152).

Freddoso, Alfred (2001) 'Suarez on Metaphysical Inquiry, Efficient Causality, and Divine Action,' personal manuscript.

Freddoso, Alfred (1988) 'Introduction' in Luis de Molina's *On Divine Foreknowledge: Part IV of the Concordia*, trans., Alfred Freddoso (Ithaca: Cornell University Press): 1-81.

Freddoso, Alfred (1986) 'The Necessity of Nature' *Midwest Studies in Philosophy* 11: 215-242.

Gallois, A. (1977) 'Van Inwagen on Free Will and Determinism' *Philosophical Studies* 32: 99-105.

Geivett, Douglas. (2002) 'Divine Providence and the Openness of God: A Response to William Hasker' *Philosophia Christi* 4: 377-398.

Ginet, Carl (2002) 'Reasons Explanations of Action: Causalist Versus Noncausalist Accounts' in Kane, ed. (2002a: 386-405).

Ginet, Carl (1997) 'Freedom, Responsibility, and Agency' *Journal of Ethics* 1: 85-98.

Ginet, Carl (1996) 'In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing' *Philosophical Perspectives* 10: 403-17.

Ginet, Carl (1995) 'Reasons Explanation of Action: An Incompatibilist Account' in O'Connor, ed. (1995a: 69-93). First printed in Ginet (1989) *Philosophical Perspectives* 3: 17-46.

Ginet, Carl (1990) *On Action* (Cambridge: Cambridge University Press).

Ginet, Carl (1983) 'In Defense of Incompatibilism' *Philosophical Studies* 44: 391-400.

Ginet, Carl (1966) 'Might We Have No Choice' in Keith Lehrer (1966): 87-104.

Goetz, Stewart. (2000) 'Naturalism and Libertarian Agency' in William Craig and J.P. Moreland, eds. (2000: 156-86).

Goetz, Stewart. (1997) 'Libertarian Choice' *Faith and Philosophy* 14: 195-211.

Goetz, Stewart. (1988) 'A Noncausal Theory of Agency' *Philosophy and Phenomenological Research* 49: 303-16.

Greenspan, Peter (1976) 'Wiggins on Historical Inevitability and Incompatibilism' *Philosophical Studies* 29: 235-247.

Harman, Gilbert (1986) 'Willing and Intending' in Richard Grandy and Richard Warner, eds., *Philosophical Grounds of Rationality: Intentions, Categories, Ends* (Oxford: Clarendon Press).

Harré, Rom and E.H. Madden (1975) *Causal Powers* (Oxford: Basil Blackwell).

Hasker, William (1999) *The Emergent Self* (Ithaca: Cornell University Press).

Hasker, William (1995) 'Middle Knowledge: A Refutation Revisited' *Faith and Philosophy* 12: 223-236.

Hasker, William (1989) *God, Time and Knowledge* (Ithaca: Cornell University Press).

Heschel, Abraham (1955) *God in Search of Man: A Philosophy of Judaism* (New York: Farrar, Straus and Giroux).

Heinaman, Robert (1986) 'Incompatibilism Without the Principle of Alternate Possibilities' *Australasian Journal of Philosophy* 64: 266-276.

Hobbes, Thomas (1969) 'Of Liberty and Necessity' in David Daiches Raphael, ed., *British Moralists, 1650-1800*, volume I (Oxford: Oxford University Press).

Hobbes, Thomas (1962) *The English Works of Thomas Hobbes*, William Molesworth, ed. (London: Scientia Aalen, 1839-45).

Holbach, Paul (1850) *System of Nature* (London: McDowell); first published in (1770), H.D. Robinson, trans. (New York: Burt Franklin).

Honderich, Ted (1996) 'Compatibilism, Incompatibilism, and the Smart Aleck' *Philosophy and Phenomenological Research* 56: 855-862.

Honderich, Ted, ed. (1995) The Oxford Companion to Philosophy (Oxford: Oxford University Press).

Honderich, Ted (1993) *How Free Are You?* (Oxford: Oxford University Press).

Honderich, Ted (1988) *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes* (Oxford: Clarendon Press).

Honderich, Ted, ed. (1973) *Essays on Freedom and Action* (London: Routledge & Kegan Paul).

Hook, Sidney, ed. (1958) *Determinism and Freedom in the Age of Modern Science* (New York: Collier-Macmillan).

Horgan, Terence (1985) 'Compatibilism and the Consequence Argument' *Philosophical Studies* 47: 339-356.

Howard-Snyder, Daniel and Jeff Jordan, eds. (1996) *Faith, Freedom, and Rationality:*

Philosophy of Religion Today (Lanham, MD: Rowman and Littlefield).

Howsepian, Avak (1999) *The Theory of Middle Freedom: A Unified Acausal Theory of Free Agency*, Ph.D. Dissertation (Notre Dame: University of Notre Dame)

Hume, David (1955) 'Of Liberty and Necessity' in his *An Enquiry Concerning Human Understanding*, ed., L.A. Selby-Bigge (Oxford: Clarendon Press). Originally published in 1743.

Hume, David (1739) *A Treatise of Human Nature* (New York: Doubleday & Co., Inc.).

Jackson, Frank (1998) *From Metaphysics to Ethics* (Oxford: Clarendon Press).

Kane, Robert, ed. (2002a) *The Oxford Handbook to Free Will* (New York: Oxford University Press).

Kane, Robert (2002b) 'Some Neglected Pathways in the Free Will Labyrinth' in Kane, ed. (2002a: 406-437).

Kane, Robert (2002c) 'Free Will: Ancient Dispute, New Themes' in Joel Feinberg and Russ Shafer-Landau, eds. (2002: 499-512).

Kane, Robert (2001) *Free Will* (Oxford: Blackwell).

Kane, Robert (2000a) 'Precis of *The Significance of Free Will*' and 'Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson' *Philosophy and Phenomenological Research* 60: 129-134, 157-167.

Kane, Robert (2000b) 'The Dual Regress of Free Will and the Role of Alternative Possibilities' *Philosophical Perspectives* 14 (Oxford: Blackwell Publishers): 57-80.

Kane, Robert (1999a) 'Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism' *Journal of Philosophy* 96: 217-240.

Kane, Robert (1999b) 'On Free Will, Responsibility and Indeterminism: Responses to Clarke, Haji and Mele' *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 2: 105-121.

Kane, Robert (1999c) 'New Directions on Free Will' *Proceedings of the 20th World Congress of Philosophy* (Boston: Boston University Press): 135-142.

Kane, Robert (1996a) *The Significance of Free Will* (New York: Oxford University Press).

Kane, Robert (1996b) 'Free Will, Responsibility and Will-Setting' *Philosophical Topics*

24: 67-90.

Kane, Robert (1995) 'Two Kinds of Incompatibilism' in O'Connor, ed. (1995a: 115-150). First printed in (1989) *Philosophy and Phenomenological Research* 50: 219-54.

Kane, Robert (1988) 'Libertarianism and Rationality Revisited' *The Southern Journal of Philosophy* 2: 441-460.

Kane, Robert (1985) *Free Will and Values* (Albany, NY: State University of New York Press).

Kapitan, Tomis (1996) 'Modal Principles in the Metaphysics of Free Will' *Philosophical Perspectives* 10: 419-445.

Kapitan, Tomis (1991a) 'Ability and Cognition: A Defense of Compatibilism' *Philosophical Studies* 63: 231-243.

Kapitan, Tomis (1991b) 'How Powerful Are We?' *American Philosophical Quarterly* 28: 331-338.

Kapitan, Tomis (1986) 'Deliberation and the Presumption of Open Alternatives' *Philosophical Quarterly* 36: 230-251.

Kenny, Anthony (1975) *Will, Freedom and Power* (Oxford: Basil Blackwell).

Kim, Jaegwon and Ernest Sosa, eds. (1999) *Metaphysics: an Anthology* (Oxford: Blackwell Publishers Ltd.).

Kim, Jaegwon (1976) 'Events as Property Exemplifications' in Myles Brand and David Walton (1976: 159-77).

Klein, Martha (1990) *Determinism, Blameworthiness, and Deprivation* (Oxford: Oxford University Press).

Kremer, Martha (2004) 'How Not to Argue for Incompatibilism,' *Erkenntnis* 60: 1-26.

Lamb, James (1993) 'Evaluative Compatibilism and the Principle of Alternate Possibilities' *Journal of Philosophy* 90: 517-527.

Lamb, James (1977) 'On a Proof of Incompatibilism' *Philosophical Review* 86: 20-35.

Lehrer, Keith (1968) 'Cans Without Ifs' *Analysis*: 29-32. Reprinted in Watson (1982: 41-5).

Lehrer, Keith, ed. (1966) *Freedom and Determinism* (New York: Random House).

Leibniz, Gottfried Wilhelm (1988) *Discourse on Metaphysics*, translated by R. Martin, D. Niall, and S. Brown (Manchester: Manchester University Press). First published in 1716.

Leibniz, Gottfried Wilhelm (1951) *Selections*, ed., P. Wiener (New York: Scribner's).

Lewis, Clive Staples (1943) *Mere Christianity* (New York, NY: Macmillan, Inc.).

Lewis, David (1999) 'New Work for a Theory of Universals' in Jaegwon Kim and Ernest Sosa, eds. (1999: 209-232). First printed in (1983) *Australasian Journal of Philosophy* 61: 343-77.

Lewis, David (1986) 'Counterfactual Dependence and Time's Arrow' in his *Philosophical Papers*, volume 2 (Oxford: Oxford University Press). First printed in (1979) *Nous* 13: 455-476.

Lewis, David (1981) 'Are We Free to Break the Laws?' *Theoria* 47: 113-121. Reprinted in Lewis (1986: 291-298).

Lewis, David (1973) *Counterfactuals* (Cambridge: Harvard University Press).

Locke, John (1894) *An Essay Concerning Human Understanding*, A.C. Fraser, ed. (Oxford: Clarendon Press). Originally published in 1689.

Lucas, John (1970) *The Freedom of the Will* (Oxford: Oxford University Press).

Lycan, William (2003) 'Free Will and the Burden of Proof' *Minds and Persons: Royal Institute of Philosophy Supplement* 53 : 107-122.

McCall, Storrs (1994) *A Model of the Universe* (Oxford: Clarendon Press).

McCann, Hugh (1998) *The Works of Agency: On Human Action, Will and Freedom* (Ithaca: Cornell University Press).

McKay, Thomas and David Johnson (1996) 'A Reconsideration of an Argument against Compatibilism' *Philosophical Topics* 24: 113-122.

McKenna, Michael (1997) 'Alternative Possibilities and the Failure of the Counterexample Strategy' *Journal of Ethics* 2: 123-142.

Melden, Al (1961) *Free Action* (London: Routledge & Kegan Paul).

Mele, Alfred (2000) 'Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance' *Philosophical Perspectives* 14: 279-300.

Mele, Alfred (1999a) 'Ultimate Responsibility and Dumb Luck' Social Philosophy and

Policy 16: 274-93.

Mele, Alfred (1999b) 'Kane, Luck, and the Significance of Free Will' *Philosophical Explorations* 2: 96-104.

Mele, Alfred (1995) *Autonomous Agents: from Self-Control to Autonomy* (New York: Oxford University Press).

Mele, Alfred (1992) *Springs of Action: Understanding Intentional Behavior* (Oxford: Oxford University Press).

Mele, Alfred and David Robb (1998) 'Rescuing Frankfurt-style Cases' *The Philosophical Review* 107: 97-112.

Merleau-Ponty, Maurice (1962) *Phenomenology of Perception*, Colin Smith, trans. (New York: Routledge).

Mill, John Stuart (1947) *On Liberty* (Oxford: Blackwell). First published in 1860.

Mill, John Stuart (1843) *A System of Logic* (New York: Harper Brothers).

Molina, Luis de (1988) *On Divine Foreknowledge: Part IV of the Concordia*, trans., Alfred Freddoso (Ithaca: Cornell University Press).

Moore, G.E. (1911) *Ethics* (London: Williams and Norgate).

Moya, Carlos (1991) *The Philosophy of Action* (Oxford: Basil Blackwell).

Nagel, Thomas (1995) 'The Problem of Autonomy' in O'Connor (1995a: 33-42). Exerpted from Nagel (1986: chapter 7).

Nagel, Thomas (1986) *The View from Nowhere* (New York: Oxford University Press).

Narveson, Jan (1977) 'Compatibilism Defended' *Philosophical Studies* 32: 83-88.

Naylor, Margery (1984) 'Frankfurt of the Principle of Alternative Possibilities' *Philosophical Studies* 46: 249-258.

Nelkin, Dana and Samuel Rickless (2002) 'Warfield's New Argument for Incompatibilism' *Analysis* 62: 104-107.

Nelkin, Dana (2001) 'The Consequence Argument and the *Mind* Argument' *Analysis* 61: 107-115.

Nozick, Robert (1995) 'Choice and Indeterminism' in O'Connor (1995: 101-114); first

published in Nozick (1981).

Nozick, Robert (1981) *Philosophical Explanations* (Cambridge, MA: Harvard University Press).

Ockham, William (1983) *Predestination, Foreknowledge, and Future Contingents*, 2nd edition, translated by Marilyn McCord Adams and Norman Kretzmann (Indianapolis: Hackett). Composed around 1335.

O'Connor, Timothy (2002a) 'The Efficacy of Reasons: a Reply to Hendrickson' *Southern Journal of Philosophy* 40: 135-137.

O'Connor, Timothy (2002b) 'Libertarian Views: Dualist and Agent-Causal Theories' in Kane, ed. (2002a: 337-355).

O'Connor, Timothy (2000a) *Person and Causes: the Metaphysics of Free Will* (New York: Oxford University Press).

O'Connor, Timothy (2000b) 'Causality, Mind, and Free Will' *Philosophical Perspectives* 14: 105-117.

O'Connor, Timothy (1996) 'Why Agent Causation?' *Philosophical Topics* 24: 143-158.

O'Connor, Timothy, ed. (1995a) *Agents, Causes, and Events* (Oxford: Oxford University Press).

O'Connor, Timothy (1995b) 'Agent Causation' in O'Connor (1995a: 173-200).

O'Connor, Timothy (1994) 'Thomas Reid on Free Agency' *Journal of the History of Philosophy* 32: 605-622.

O'Connor, Timothy (1993a) 'Indeterminism and Free Agency: Three Recent Views' *Philosophy and Phenomenological Research* 53: 499-526.

O'Connor, Timothy (1993b) 'On the Transfer of Necessity' *Nous* 27: 204-218.

Otsuka, Michael (1998) 'Incompatibilism and the Avoidability of Blame' *Ethics* 108: 685-701.

Otte, Richard (1987) 'A Defense of Middle Knowledge' *International Journal for Philosophy of Religion* 21: 161-169.

Papineau, David (1995) 'Laws, Natural or Scientific' in Honderich, ed. (1995: 474-476).

Peacocke, Christopher (1999) *Being Known* (Oxford: Clarendon Press).

Pereboom, Derk (2001) *Living without Free Will* (Cambridge: Cambridge University Press).

Pereboom, Derk (2000) 'Alternative Possibilities and Causal Histories' *Philosophical Perspectives* 14: 119-137.

Pereboom, Derk (1996) 'Soft Libertarianism and Frankfurt-style Scenarios' *Philosophical Topics* 24: 123-141.

Pettit, Gordon (2000) *Conditions of Moral Responsibility*, Ph.D. Dissertation (Notre Dame: University of Notre Dame).

Plantinga, Alvin (2000) *Warranted Christian Belief* (Oxford: Oxford University Press).

Plantinga, Alvin (1993a) *Warrant: the Current Debate* (Oxford: Oxford University Press).

Plantinga, Alvin (1993b) *Warrant and Proper Function* (Oxford: Oxford University Press).

Plantinga, Alvin (1990) 'Advice to Christian Philosophers' in Michael Beaty, ed., *Christian Theism and the Problems of Philosophy* (Notre Dame: University of Notre Dame Press). Previously published in (1984: 253-271) *Faith and Philosophy* 1.

Plantinga, Alvin (1985) 'Replies' in *Alvin Plantinga*, eds., James Tomberlin and Peter van Inwagen (Dordrecht: D. Reidel)

Plantinga, Alvin (1977) *God, Freedom, and Evil* (Grand Rapids, MI: Eerdmans).

Plantinga, Alvin (1974) *The Nature of Necessity* (Oxford: Clarendon Press).

Ratzsch, Del (1990) 'Nomo(theo)logical Necessity' in Michael Beaty, ed. (1990). First printed in (1987: 383-402) *Faith and Philosophy* 4.

Ravizza, Michael (1994) 'Semi-Compatibilism and the Transfer of Nonresponsibility' *Philosophical Studies* 75: 61-94.

Reid, Thomas (2001/1792) 'Of Power' *Philosophical Quarterly*: 3-12.

Reid, Thomas (1895) *Essays on the Active Powers of the Human Mind* in, ed., Sir William Hamilton (1983) *The Works of Thomas Reid, D.D.*, 8th edition, (Edinburgh: Maclachlan and Stewart).

Reid, Thomas (1983) 'Reid's Unpublished Letters' *The Works of Thomas Reid, D.D.*,

8th edition, Sir William Hamilton, ed. (Edinburgh: Maclachlan and Stewart): 38-92. First published between 1763 and 1793.

Rowe, William (2000) 'The Metaphysics of Freedom: Reid's Theory of Agent Causation' *American Catholic Philosophical Quarterly*: 425-446.

Rowe, William (1995) 'Two Concepts of Freedom' in O'Connor (1995a). First printed a Presidential Address: *Proceedings of the American Philosophical Association* 62: 43-64.

Rowe, William (1991a) *Thomas Reid on Freedom and Morality* (Ithaca: Cornell University Press).

Rowe, William (1991b) 'Responsibility, Agent Causation, and Freedom: an 18th Century' *Ethics* 101: 237-257.

Rowe, William (1989) 'Causing and Being Responsible for what is Inevitable' *American Philosophical Quarterly*: 153-9.

Rowe, William (1987) 'Two Concepts of Freedom' *Proceedings and Addresses of the American Philosophical Association* 61: 43-64.

Rowe, William (1982) 'Two Criticisms of the Agency Theory' *Philosophical Studies* 42: 363-378.

Russell, Bertrand (1953) 'On the Notion of Cause, with Applications to the Free-Will Problem' *Readings in the Philosophy of Science*, ed., Herbert F. and M. Brodbeck (New York: Appleton-Century-Crofts)

Russell, Paul (1984) 'Sorabji and the Dilemma of Determinism' *Analysis* 44: 166-72.

Sedley, David and Anthony Long (1999) *The Hellenistic Philosophers, Volume I, Translations and Commentary* (Cambridge: Cambridge University Press). First edition printed (1987).

Schlick, Moritz (1939) *Problems of Ethics*, translated by David Rynin (New York: Prentice Hall).

Searle, John (1984) *Minds, Brains, and Science* (Cambridge, Massachusetts: Harvard University Press).

Searle, John (1983) *Intentionality* (Cambridge: Cambridge University Press).

Shoemaker, Sidney (1998) 'Causal and Metaphysical Necessity' *Pacific Philosophical Quarterly* 79: 59-77.

Shoemaker, Sidney (1980) 'Causality and Properties' in van Inwagen, ed. (1980: 109-35).

Slote, Michael (1982) 'Selective Necessity and the Free Will Problem' *Journal of Philosophy* 79: 5-24.

Sobel, Jordan (1998) *Puzzles for the Will: Fatalism, Newcomb and Samarra, Determinism and Omniscience* (Toronto: University of Toronto Press).

Sorabji, Richard (1980) *Necessity, Cause and Blame: Perspectives on Aristotle's Philosophy* (Ithaca: Cornell University Press).

Sosa, Ernest, ed. (1979) *Essays on the Philosophy of Roderick Chisholm* (Amsterdam: Rodopi).

Spinoza, Benedict (1677) *Ethics*, translated by W.H. White and A.H. Sterling (Oxford: Oxford University Press, 1927).

Strawson, Galen (1995) 'Libertarianism, Action, and Self-Determination' in O'Connor (1995a: 13-31). Exerpted from Strawson (1986: chapter 2).

Strawson, Galen (1986) *Freedom and Belief* (Oxford: Clarendon Press).

Strawson, Peter (1962) 'Freedom and Resentment' Proceedings of the British Academy' 48: 1-25. Reprinted in Watson, ed. (1982: 59-80).

Stump, Eleanore (1996) 'Libertarian Freedom and the Principle of Alternative Possibilities' in Daniel Howard-Snyder and Jeff Jordan, eds. (1996: 73-88).

Stump, Eleanore (1995) 'Intellect, Will, and the Principle of Alternate Possibilities' in Michael Beaty, ed. (1995: 254-285).

Stump, Eleanore and John Fischer (2000) 'Transfer Principles and Moral Responsibility' *Philosophical Perspectives* 14: 47-55.

Suarez, Francisco (1994) *On Efficient Causality: Metaphysical Disputations 17, 18, and 19*, Alfred Freddoso, trans. (New Haven, CT: Yale University Press). First printed (1597).

Suppe, Frederick (1977) *The Structure of Scientific Theories* (Urbana: University of Illinois Press).

Swoyer, Chris (1982) 'The Nature of Natural Laws' *Australasian Journal of Philosophy* 60: 203-23.

Taylor, Richard (1992) *Action and Purpose*, 4th edition (Englewood Cliffs, NJ:

Prentice-Hall). First printed in (1963).

Thalberg, Irving (1983) *Misconceptions of Mind and Freedom* (New York: University Press of America).

Thalberg, Irving (1976) 'How Does Agent Causality Work?' in Myles Brand and David Walton, eds. (1976: 213-238).

Tomberlin, James, ed. (1989) *Philosophical Perspectives, 3, Philosophy of Mind and Action Theory* (Atascadero, CA: Ridgeview Publishing Co.).

Tooley, Michael (1987) *Causation: A Realist Approach* (Oxford: Oxford University Press).

Tooley, Michael (1977) 'The Nature of a Law' *Canadian Journal of Philosophy* 7: 667-98.

Thorp, John (1980) *Free Will: A Defense Against Neurophysiological Determinism* (London: Routledge & Kegan Paul).

Van Fraassen, Bas (1989) *Laws and Symmetries* (Oxford: Clarendon Press).

Van Inwagen, Peter (2001) 'Van Inwagen on Free Will' *Inland Northwest Philosophy Conference*, manuscript.

Van Inwagen, Peter (2000) 'Free Will Remains a Mystery' *Philosophical Perspectives* 14: 1-19.

Van Inwagen, Peter (1999) 'Moral Responsibility, Determinism, and the Ability To Do Otherwise' *The Journal of Ethics* 3: 341-350.

Van Inwagen, Peter (1997) 'Against Middle Knowledge' *Midwest Studies in Philosophy* 21: 225-36.

Van Inwagen, Peter (1995a) 'When is the Will Free?' in O'Connor (1995a: 219-238). First printed in James Tomberlin, ed. (1989).

Van Inwagen, Peter (1995b) 'The Place of Chance in a World Sustained by God' in his *God Knowledge and Mystery: Essays in Philosophical Theology* (Ithaca, NY: Cornell University Press): 42-65. Reprinted from Thomas Morris, ed. (1988) *Divine and Human Action: Essays in the Metaphysics of Theism* (Ithaca, NY: Cornell University Press).

Van Inwagen, Peter (1994) 'When Is the Will Not Free?' *Philosophical Studies* 75: 95-114.

Van Inwagen, Peter (1993) *Metaphysics* (Boulder, CO: Westview Press, Inc.)

Van Inwagen, Peter (1990a) *Material Beings* (Ithaca: Cornell University Press).

Van Inwagen, Peter (1990b) 'Logic and the Free Will Problem' *Social Theory and Practice* 16: 277-290.

Van Inwagen, Peter (1988) 'The Magnitude, Duration, and Distribution of Evil: A Theodicy' *Philosophical Topis* 16: 161-187.

Van Inwagen, Peter (1986a) 'The Incompatibility of Responsibility and Determinism' in Fischer (1986: 241-249). First printed in Michael Bradie and Myles Brand (1980: 30-37).

Van Inwagen, Peter (1986b) 'Ability and Responsibility' in John Fischer, ed. (1986a: 153-173). First printed in (1978) *Philosophical Review* 87: 201-224.

Van Inwagen, Peter (1983) *An Essay on Free Will* (Oxford: Clarendon Press).

Van Inwagen, Peter, ed. (1980) *Time and Cause* (Dordrecht: Reidel).

Van Inwagen, Peter (1978) 'Ability and Responsibility' *Philosophical Review* 87: 201-224.

Van Inwagen, Peter (1977) 'Reply to Narveson' *Philosophical Studies* 32: 89-98. Earman, J. (1986) *A Primer on Determinism* (Dordrecht: D. Reidel).

Van Inwagen, Peter (1975) 'The Incompatibility of Free Will and Determinism' *Philosophical Studies* 27: 185-99. Reprinted in Watson (1982: 46-58).

Van Inwagen, Peter (1974) 'A Formal Approach to the Problem of Free Will and Determinism' *Theoria* 24 (Part I): 9-22.

Van Inwagen, Peter and Dean Zimmerman, eds. (1998) *Metaphysics: The Big Questions* (Massachusetts: Blackwell Publishers).

Vihvelin, Kadri (1995a) 'Causes, Effects, and Counterfactual Dependence' *Australasian Journal of Philosophy* 73: 560-573.

Vihvelin, Kadri (1995b) 'Reply to "Causes and Nested Conditionals"' *Australasian Journal of Philosophy* 73: 579-581.

Vihvelin, Kadri (1991) 'Freedom, Causation, and Counterfactuals' *Philosophical Studies* 64: 161-184. Waller, Bruce (1990) *Freedom Without Responsibility* (Philadelphia: Temple University Press).

Warfield, Ted (2003) 'Compatibilism and Incompatibilism: Some Arguments' in

Michael Loux and Dean Zimmerman, eds. (2003: 613-631), *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press).

Warfield, Ted (2000) 'Causal Determinism and Human Freedom are Incompatible: a New Argument for Incompatibilism' *Philosophical Perspectives: Action and Freedom* 14: 167-180.

Warfield, Ted (1999) 'Donald Davidson's Freedom' *Interpretations and Causes: New Perspectives on Donald Davidson's Philosophy*, Mario De Caro, ed. (Netherlands: Kluwer Academic Publishers).

Warfield, Ted (1996) 'Determinism and Moral Responsibility Are Incompatible' *Philosophical Topics* 24: 215-226.

Watson, Gary (1987) 'Free Action and Free Will' *Mind* 96: 145-72.

Watson, Gary, ed. (1982) *Free Will* (New York: Oxford University Press).

Weinstock, Jerome (1976) 'Reid's Definition of Freedom' in Stephen Barker and Tom Beauchamp, eds. (1976: 95-102).

Widerker, David (2000) 'Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look' *Philosophical Perspectives* 14: 181-201.

Widerker, David (1995a) 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities' *The Philosophical Review* 104: 247-261.

Widerker, David (1995b) 'Libertarian Freedom and the Avoidability of Decisions' *Faith and Philosophy* 12: 113-118.

Widerker, David (1991) 'Frankfurt on "Ought" Implies "Can" and Alternate Possibilities' *Analysis* 49: 222-224.

Widerker, David (1987) 'On an Argument for Incompatibilism' *Analysis* 47: 37-41.

Wiggins, David (1973) 'Towards a Reasonable Libertarianism' in Ted Honderich, ed. (1973: 31-61).

Wierenga, Edward (2001) 'Providence, Middle Knowledge, and the Grounding Objection' *Philosophia Christi* 3: 447-458.

Wierenga, Edward (1991) 'Prophecy, Freedom, and the Necessity of the Past' *Philosophical Perspectives* 5: 425-446.

Wierenga, Edward (1989) *The Nature of God* (Ithaca: Cornell University Press).

Wolf, Susan (1990) *Freedom within Reason* (Oxford: Oxford University Press).

Wolfram, Stephen (2002) *A New Kind of Science* (Champaign, IL: Wolfram Media, Inc.)

Wyma, Keith (1997) 'Moral Responsibility and the Leeway for Action' *American Philosophical Quarterly* 34: 57-70.

Zagzebski, Linda (2000) 'Does Libertarian Freedom Require Alternate Possibilities?' *Philosophical Perspectives* 14: 231-248.

Zimmerman, Michael (1988) *An Essay on Moral Responsibility* (New Jersey: Rowman & Littlefield).