

THAT WE DO NOT HAVE FREE WILL: A CRITIQUE OF LIBERTARIANISM

By

KIRK W. WALKER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

© 2012 Kirk W. Walker

To my Mom

ACKNOWLEDGMENTS

I thank Professor Robert D'Amico for his guidance and encouragement over the years. I thank Professors John Biro, Chuang Liu, and Frederick Gregory for their helpful suggestions. And I thank the Philosophy Department for its financial support.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
ABSTRACT.....	7
CHAPTER	
1 ARGUMENTS FOR INCOMPATIBILISM.....	9
Introduction.....	9
Van Inwagen’s Consequence Argument.....	10
Lewis’ Objection.....	11
The Principle of Agglomeration.....	15
The Revised Consequence Argument.....	16
The Irrelevance of Agglomeration.....	17
The Zygote Argument.....	19
Objections and Replies.....	20
Conclusion.....	24
2 AGENT-CAUSATION.....	26
The Basic Account—Chisholm.....	26
Objections to the Basic Account.....	29
Agent Causation and Reasons Explanation—O’Connor.....	32
Further Objections.....	35
A Hybrid Account—Clarke.....	38
Objections to Clarke’s Account.....	39
Metaphysical Considerations—O’Connor.....	42
Conclusion.....	46
3 NONCAUSALIST LIBERTARIANISM.....	47
The Rejection of Causalism.....	48
Objections.....	49
Noncausalist Phenomenology of Action.....	51
Objections.....	53
Noncausalist Metaphysics of Action.....	54
Objections.....	55
Noncausalist Explanation of Action.....	59
Objections and Discussion.....	60
Conclusion.....	66

4	EVENT-CAUSAL LIBERTARIANISM	67
	Introduction	67
	Early Event-Causal Theories	67
	Van Inwagen's Account	74
	Recent Event-Causal Theories	78
	Daring Soft Libertarianism	83
5	KANE'S EVENT-CAUSAL LIBERTARIANISM.....	90
	The Elements of Kane's Theory.....	90
	Objections to Kane's Theory	95
	Conclusion	107
6	POSTSCRIPT	108
	The Treatment of Criminals	108
	Attitude Towards Life	110
	Inter-Personal Relationships.....	112
	How Things Would Change	114
	Conclusion	117
	REFERENCE LIST.....	118
	BIOGRAPHICAL SKETCH.....	123

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

THAT WE DO NOT HAVE FREE WILL: A CRITIQUE OF LIBERTARIANISM

By

Kirk W. Walker

December 2012

Chair: Robert D'Amico
Major: Philosophy

My dissertation concerns the understanding of the notion of free will and whether or not our prima facie belief that we possess it is justified. Specifically, I will argue that our belief that we possess free will is not justified. My overall strategy will be an argument by process of elimination to the effect that once the most promising theories and arguments in favor of free will are eliminated, the rational conclusion to reach is that we do not have free will.

The dissertation has six chapters, two devoted to general argumentation and overview, and four devoted to incisive analysis. My first task is to examine arguments against compatibilism, which then motivates my examination of the incompatibilist libertarian theories.

The *event-causal theories* comprise one family of libertarian theory. These theories argue that under the right conditions, certain agent-involving events nondeterministically cause the free actions of a human agent. Typically, such theories take advantage of the general conditions for a free action that are specified by compatibilist theories, and then add an indeterministic causal factor. *Noncausal* accounts comprise another family of libertarian theories. These theories are notable in

holding that free actions are neither caused by anything nor have any internal causal structure. Typically, they maintain that a basic action, or basic mental action, has an intrinsic quality that is characteristic of free actions. The final family of libertarianism comprises the *agent-causal* theories. According to these theories, human beings possess a unique kind of causal power not found in the rest of nature. On these accounts, the agent, as a persisting substance, has the power to initiate causal chains independently of any event whatsoever, including the reasons of the agent.

I dedicate one chapter apiece to showing why each family of libertarian theory is in error. In addition I devote a chapter to criticizing the event-causal theory of Robert Kane. In the conclusion I take stock of where the critique of libertarianism leaves us.

CHAPTER 1 ARGUMENTS FOR INCOMPATIBILISM

Introduction

The overarching aim of the present dissertation is to make a strong case for the *nonexistence* of free will. To this end, the bulk of the text will be devoted to showing why the leading libertarian theories are in error. All *proponents* of free will contend that natural laws and human nature are both such as to make free will either possible or actual. Libertarians argue that the pattern of cause and effect in nature is not entirely deterministic.¹ Libertarians argue that one factor or another, depending on the particular theory, permits human agency to depart from otherwise universal determinism. But libertarianism and the no-free-will thesis are not the only available alternatives. Libertarians are to be distinguished from another group of free will proponents, the compatibilists. Compatibilists allow the existence of universal determinism, yet insist this state of affairs is compatible with a correct understanding of what it means to be free.

Because the literature on free will is vast, I will not be able to survey compatibilism in depth. The aim of my dissertation is provide an account of the major incompatibilist free will positions that are available. I show that these positions face serious and in some cases even insurmountable problems.

As regards the position of compatibilism (in which we are able to maintain our concept of freedom of action while still adhering to determinism) while I cover only some of the debate about that view in my first chapter, my aim is more modest. Given the

¹ Determinism is the thesis that every event, including every human event, act, and decision is the inevitable consequence of antecedent states of affairs.

constraints of time and length, I cannot claim to entirely set aside this position nor show that the position can be successfully blocked. But I do think it is important to show the reader why the compatibilist position faces such serious challenges that one might, in seeing the nature of these challenges, be motivated to explore the incompatibilist route with regard to the debate.

Van Inwagen's Consequence Argument

The most widely discussed argument for incompatibilism is the so-called Consequence Argument articulated by Peter van Inwagen in *An Essay on Free Will* (1983, pp. 56-82). Roughly, the argument says that if determinism is true and one has neither a choice about the laws of nature nor the state of the world in the distant past, then one has no choice about any of one's actions. In other words, if our present (and future) actions are the deterministic consequences of the combination of the laws of nature and the configuration of the universe at some time before there were any agents, then our inability to affect the past and the laws of nature transfers to our present (and future) actions so that they are the inevitable consequences of what went before. Thus given two plausible assumptions concerning our relation to the past and the laws of nature, the presence of determinism appears to rule out the possibility of free will. Intuitively, this is a very powerful argument.

In order to better appraise the force of this argument, let us turn to the formal articulation of it:

Let us read 'Np' as 'p and no one has or ever had any choice about whether p'. We employ the following two inference rules

$\alpha \quad \Box p \vdash Np$

$\beta \quad Np, N(p \rightarrow Nq) \vdash Nq$

(The box, of course, represents necessity or truth in all possible worlds.) Let 'L' represent the conjunction of the laws of nature into a single proposition. Let 'P₀' represent the proposition that describes the state of the world at some time in the remote past. Let 'P' represent any true proposition. The following statement, proposition (1), is a consequence of determinism:

$$(1) \quad \Box((P_0 \ \& \ L) \rightarrow P).$$

I now argue,

$$(2) \quad \Box(P_0 \rightarrow [L \rightarrow P]) \quad 1, \text{ standard logic}$$

$$(3) \quad N(P_0 \rightarrow [L \rightarrow P]) \quad 2, \quad \alpha$$

$$(4) \quad NP_0 \quad \text{Premise}$$

$$(5) \quad N(L \rightarrow P) \quad 3, 4 \ \beta$$

$$(6) \quad NL \quad \text{Premise}$$

$$(7) \quad NP \quad 5, 6, \ \beta$$

Since the two premises are obviously true—no one has any choice about the past; no one has any choice about the laws of nature—, (7) follows from (1) if the two rules of inference are valid. And from this it follows that if determinism is true, no one has any choice about anything. (van Inwagen 2000, p. 159)

In the above argument, the two premises seem unassailable. Similarly, inference rule alpha seems beyond reproach as no one has any choice about the truth or falsity of a necessarily true proposition. This leaves beta as a possible weakness in the argument; and as we shall see below critics have attacked the argument at this point. However, before taking a closer look at beta, let us examine one of the earliest criticisms of the Consequence Argument.

Lewis' Objection

David Lewis argued that step (7) does not follow from step (6) for the reason that the ability to break a law of nature does not imply a god-like ability outside the reach of

human agents. In short, he subjects NL to scrutiny not because it is false on the face of it, but because one than one thing may be meant by denying it. Lewis first summarizes the scenario specified by the consequence argument:

I have just put my hand down on my desk. That, let me claim, was a free but predetermined act. I was able to act otherwise, for instance to raise my hand. But there is a true historical proposition H about the intrinsic state of the world long ago, and there is a true proposition L specifying the laws of nature that govern our world, such that H and L jointly determine what I did. They jointly imply the proposition that I put my hand down. They jointly contradict the proposition that I raised my hand. Yet I was free; I was able to raise my hand. The way in which I was determined not to was not the sort of way that counts as inability.

What if I had raised my hand? Then at least one of three things would have been true. Contradictions would have been true together; or the historical proposition H would not have been true; or the law proposition L would not have been true. (Lewis 1981, pp. 122-123)

This fairly represents the situation posed by the Consequence argument. The compatibilist seems to be face with three impossibilities, namely that two propositions that jointly contradict themselves would both be true, that the state of the universe was not as it was, and/or the agent is capable of violating the laws of nature. However, Lewis' strategy is to argue that the Consequence argument equivocates in its claim that NL, no one has a choice about the laws of nature. He distinguishes between two possible theses that may be implied by the denial of NL (Lewis 1981, p. 123). The stronger of the two theses states that "I am able to break a law." This would indeed be a god-like power as it seems preposterous that an ordinary entity can act in a manner contrary to the laws of nature. But Lewis does not think the compatibilist is committed to this thesis. Instead, the denier of NL may intend a weaker thesis: "I am able to do something such that, if I did it, a law would be broken." This alternative is far more plausible than the first and Lewis thinks it is the one the compatibilist should be seen as

asserting. Moreover, if this what the denial of NL means, then the Consequence Argument may not preclude an agent acting in a manner that in one sense appears ruled out by the deterministic consequences of the state of the world in the distant past and the totality of the laws of nature.

Lewis elaborates as follows:

Now consider the disputed case. I am able to raise my hand, although it is predetermined that I will not. If I raised my hand, some law would be broken. I even grant that a law-breaking event would take place. (Here I use the present tense neutrally. I mean to imply nothing about when a law-breaking event would take place.) But is it so that my act of raising my hand would cause any law-breaking event? Is it so that my act of raising my hand would itself be a law-breaking event? Is it so that any other act of mine would cause or would be a law-breaking event? If not, then my ability to raise my hand confers no marvelous ability to break a law, even though a law would be broken if I did it.

Had I raised my hand, a law would have been broken beforehand. The course of events would have diverged from the actual course of events a little while before I raised my hand, and at the point of divergence there would have been a law-breaking event — a divergence miracle, as I have called it. But this divergence miracle would not have been caused by my raising my hand. If anything, the causation would have been the other way around. Nor would the divergence miracle have been my act of raising my hand. That act was altogether absent from the actual course of events, so it cannot get under way until there is already some divergence. Nor would it have been caused by any other act of mine, earlier or later. Nor would it have been any other act of mine. Nor is there any reason to say that if I had raised my hand there would have been some other law-breaking event besides the divergence miracle; still less, that some other law-breaking event would have been caused by, or would have been, my act of raising my hand. To accommodate my hypothetical raising of my hand while holding fixed all that can and should be held fixed, it is necessary to suppose one divergence miracle, gratuitous to suppose any further law-breaking.

Thus I insist that I was able to raise my hand, and I acknowledge that a law would have been broken had I done so, but I deny that I am therefore able to break a law. To uphold my instance of soft determinism, I need not claim any incredible powers. To uphold the compatibilism that I actually believe, I need not claim that such powers are even possible. (Lewis 1981, pp. 124-125)

Lewis argues that in the scenario where the state of the universe and the laws of nature jointly determine that he is unable to raise his hand, he would nonetheless be able to. But, he goes on to say, one should not conclude from this that the compatibilist attributes to him the remarkable power to break a law of nature. This is because the raising of his hand neither is nor is the cause of a law-breaking event. Rather, the law would be broken some time before the hand is raised. The law-breaking event is independent of the actual raising of the hand. This prior, independent breaking of the law is indeed a miracle, a divergence miracle. But it does not follow that this miracle applies to the hand raising. Hence a godlike power to miraculously break the laws of nature is not assumed.

Although Lewis marks a distinction between the power to break a law and the power to do something on the occasion that a law is broken, there is really no practical difference between the two on the assumption that the agent has the ability to raise his hand. In the case the compatibilist rejects, a person directly breaks one or more laws of nature by either the action being a law-breaking event or by it causing a law-breaking event. Under this scenario, the agent evidently has the godlike ability to act in a manner ruled out by the laws of nature. This is clearly unbelievable. As we can see, David Lewis's argument shows that there is a way to have the premises of the consequence argument concerning the fixity of the laws of nature and the events of the past (there is nothing anyone can do about them), while still retaining the notion of there being an alternative action available to the agent. Lewis argues that such a counter-factual option for the agent can be accommodated by proposing that in such a case, if the agent were to take that alternate option, then the laws would have to be broken; there would then

be a “miraculous” break with the laws and the past. Now we can leave his argument as such because it only adds support to my case. It shows that it is very difficult for the compatibilist position to retain anything like the intuitive notion of alternative possibilities that motivates compatibilism in the first place. Thus the way in which Lewis salvages compatibilism is in my view a further reason to doubt the plausibility of the position. But since there are compatibilists who reject the principle of alternative possibilities, this dissertation is not the occasion to try to definitively resolve the matter.

The Principle of Agglomeration

As previously mentioned, the inference rule beta has been subjected to scrutiny. As it turns out, McKay and Johnson have proven that beta entails agglomeration, which in turn is invalid. First, agglomeration is derived from Np and Nq by means of rules alpha and beta.

Principle of Agglomeration: $Np, Nq \vdash N(p \& q)$	
1.) Np	Premise
2.) Nq	Premise
3.) $\Box [p \rightarrow (q \rightarrow (p \& q))]$	Necessity of a logical truth
4.) $N[p \rightarrow (q \rightarrow (p \& q))]$	From 3 and α
5.) $N[q \rightarrow (p \& q)]$	From 1, 4, and β
6.) $N(p \& q)$	From 2, 5 and β

Since rule alpha is beyond suspicion, a counterexample to agglomeration will prove that beta is invalid.

McKay and Johnson’s counterexample unfolds as follows (McKay and Johnson, 1996, p. 115). Suppose I have a fair coin that I could have tossed yesterday, but didn’t. Assume that no one else could have tossed it. Also assume that if I had tossed the coin it would either land heads or tails. That is, we are to exclude the possibility of it landing on its edge or a bird scooping it out of the air, etc. Then both of the following are true:

N The coin did not land "heads" yesterday.
N The coin did not land "tails" yesterday.

These propositions are true since it would have been a matter of chance whether the coin had landed heads or tails, and I would have had no choice over what is a matter of chance. But now consider the conjunctive proposition:

N (The coin did not land "heads" yesterday & the coin did not land "tails" yesterday).

This proposition is false. For although *as it stands* the conjunction is true, as I did not toss the coin, I *could have* made the conjunction false by tossing the coin. If I had tossed the coin one of the conjuncts would have come out true, and the other false. But then by the rules of sentential logic, the overall conjunction would have been false. And that was something I had a choice about since it was within my power to toss the coin. This counterexample shows that rule beta is invalid.

The Revised Consequence Argument

Although in its original formulation the Consequence Argument fails Finch and Warfield have devised a way to repair it. They replace rule beta with rule *beta 2*. As in van Inwagen's original argument, ' Np ' is the abbreviation for " p and no one has, or ever had, any choice about whether p " and ' \square ' stands for broad logical necessity.

Rule Beta 2

$\beta 2) \quad (Np \ \& \ \square (p \rightarrow q)) \rightarrow Nq$ (Finch and Warfield 1998, p. 521)

Beta 2 merely claims that one has no choice about the logical consequences of those things one has no choice about. Next they formulate a streamlined version of the original argument using the new rule of inference. In the argument determinism is to be understood as the thesis that the conjunction of the past and the laws of nature fixed a unique future. P is a proposition expressing the entire state of the world at a time in the

distant past, before there were any agents. L is the conjunction of all the laws of nature.

And F stands for any truth.

The Improved Consequence Argument

P1. $\Box\{(P \ \& \ L) \rightarrow F\}$ Consequence of determinism

P2. $N(P \ \& \ L)$ Fixity of the past and laws

C1. NF P1, P2, Beta 2

(Finch and Warfield 1998, p. 522)

The argument says we are to suppose that any given truth follows by logical necessity from the conjunction of the past and the laws of nature. Next we are to assume that no one has, or ever had, any choice about the conjunction of the past and the laws of nature. Finally, using beta 2 we infer that no one has a choice about the logical consequence, F, of those things no one has a choice about. Since agglomeration cannot be derived from beta 2, the Improved Consequence Argument gives us compelling reasons for embracing incompatibilism (and rejecting compatibilism). This is a considerable advance in the debate between incompatibilists and compatibilists, but, as we shall shortly see, it may be unnecessary.

The Irrelevance of Agglomeration.

Crisp and Warfield propose two desiderata or prerequisites, to the question of the validity of rule β : (1) Proposed counterexamples to principle β must not presuppose the compatibility of freedom and determinism; and (2) Proposed counterexamples to principle β must not presuppose the truth of indeterminism. (Crisp and Warfield 2000 pp. 175-180) They argue that the objector to principle Beta is placed in a quandry:

As we see it, McKay and Johnson face a dilemma. Either their coin flip example occurs in a deterministic world or it occurs in an indeterministic world. If the latter, then, as McKay and Johnson point out, they would be vulnerable to the introduction of the Delta principle. So, to meet the second desideratum, the world in question must be deterministic. But if so, they violate the first desideratum. For recall how McKay and Johnson introduce their coin flip scenario: "Suppose that I do not toss a coin but

could have" (1996, p. 118). If the world in question is deterministic, however, this amounts to assuming that one could have done other than what one did in a deterministic world. But this implies compatibilism and thus violates the first desideratum on proposed counterexamples to Principle Beta. So, while McKay and Johnson have a strong argument against Principle Beta, their argument should be of little concern to proponents of the Consequence argument. (Crisp and Warfield 2000 p. 180)

The principle Delta referred to in the above passage is

Delta: $D, Np, N(p \rightarrow q) \vdash Nq$

where D stands for the thesis of determinism. The weak spot of the Consequence Argument was seen to be rule beta. Once agglomeration was derived by means of both alpha and beta the conclusion we were led to is that the latter is invalid. But Crisp and Warfield now argue that this objection to the Consequence Argument runs afoul of the two desiderata. On the one hand, the objection begs the question against the incompatibilist by invoking the compatibility of determinism and free will. It does this by saying that while the agent could have flipped the coin, she didn't. This presupposes that one could have done otherwise than what one is determined to do and is thus a compatibilist assumption. But suppose on the other hand the objection takes indeterminism as given. This does not worry the defender of the Consequence Argument who can simply introduce principle delta which incorporates determinism. After all, the main point of the argument is that free will is incompatible with determinism; hence it is perfectly alright to incorporate determinism into its premises. And with the addition of principle delta, the revised argument of Crisp and Warfield thus establishes that the critique of rule beta fails to undermine the Consequence Argument. In summary, it is considered whether the coin toss is a deterministic process or an indeterministic one. If it is the latter, then the simple inclusion of determinism precludes

the threat to the Consequence argument. However, if it is the former, then the objection to the Consequence Argument would have a false premise by violating the first desiderata. With the failure of the critique of rule beta, the Consequence Argument thus continues to be highly persuasive in establishing that incompatibilism is true, that free will and determinism are not compatible.

The Zygote Argument

In *Free Will and Luck* Alfred Mele presents, without endorsing, a new argument for incompatibilism. He first describes a scenario in which a zygote is produced by extraordinary means.

Diana [a goddess with special powers] creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in May will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E. If this agent, Ernie, has any unsheddable values at the time, they play no role in motivating his A-ing. Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises

his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. So he satisfies a version of my proposed compatibilist sufficient conditions for having freely A-d. (Mele 2006 p. 188)

Mele then proceeds to delineate the zygote argument proper:

- (1) Because of the way his zygote was produced in his deterministic universe, Ernie is not a free agent and is not morally responsible for anything.
- (2) Concerning free action and moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal human zygote comes to exist in a deterministic universe.

(3) So determinism precludes free action and moral responsibility. (Mele 2006 p. 189)

Mele offers as a reason for accepting premise one that the zygote in question resembles a robot that has been programmed to respond in specific ways to inputs or stimuli. This follows from Diana apparently building into Ernie from the outset, all the characters, including patterns of reasoning, that will lead Ernie to A at the desired time. Despite the fulfillment of the compatibilist conditions, the presence of an overseer who designs every detail of the agent's development and final state serves to neutralize the liberating impact of those conditions with the result that the agent is as unfree and as pre-programmed as a robot. Mele's reason for accepting premise two is that with respect to deterministic universes, there is no difference between a zygote produced by an intelligence, and one produced by blind forces. Consider deterministic universe U^* which is just like universe U , except that in U^* Ernie is created by blindly operating forces. The blind forces impose a pattern with as much thoroughness and certainty as if there were a conscious designer at work. Moreover, the resultant constitution of the agent would be the same whether it was created by a designer or by blind forces.

Objections and Replies

John Martin Fischer has criticized premise one of the zygote argument by comparing the goddess Diana to a couple who produce the zygote by normal means.

Now we have a puzzle. If we start with the John and Mary scenario depicted above, we get the conclusion that the mental states and intentions of the distal creators of the zygote are irrelevant to Ernie's subsequent moral responsibility. And, by the way, it surely cannot matter if a couple intends to create a zygote with certain properties, as opposed to an individual, holding everything else fixed. Thus one could substitute Diana for John and Mary, and the clear intuition – proceeding in this way – would be that Diana's intentions are irrelevant to Ernie's moral responsibility 30 years down the road. But if we start with Diana, we can

come to quite the opposite conclusion, as we are invited to do by Premiss 1 of the Zygote Argument.

I thus contend that the Diana scenario cannot in itself help to establish Premiss 1 of the Zygote Argument. After all, one might as well consider the John and Mary scenario; nothing in the Zygote Argument entails or even suggests that somehow starting with the Diana scenario is more plausible than – or preferable to – starting with the John and Mary scenario. But the John and Mary scenario pulls in quite the opposite direction from the Diana scenario; indeed, it points to the rejection of 1. Thus the Zygote Argument is unsuccessful; or, at the least, the Diana scenario does not provide strong reason to accept Premiss 1. That is, in light of the John and Mary scenario, the support offered for Premiss 1 by the Diana scenario is considerably weakened. (Fischer 211, pp. 269-270)

In Fischer's scenario, John and Mary create a zygote thanks to the unforeseen failure of their contraception. Fischer maintains that since the mental states and intentions of John and Mary are irrelevant to Ernie's subsequent moral responsibility, we should conclude that Diana's mental states and intentions are also irrelevant to Ernie's subsequent moral responsibility. He argues that whatever intuitive plausibility the Diana scenario has derives from our starting with it. But if we start with an ordinary couple, premise one is no longer persuasive.

Surely Fischer is wrong about this. The reason for starting with Diana is because she is a goddess capable of designing a zygote with just the configuration it needs to have so that deterministic forces will lead it to A on a given occasion at a particular time in the future. The fact that Diana is a goddess with special powers is crucial to the scenario. An ordinary couple cannot specify the initial configuration of a zygote so that it with certainty will behave in a specific way at a particular time once it has developed. Likewise, an ordinary couple cannot anticipate the deterministic operations of the universe over the course of many years in order to fine tune a given result as subtle as a given individual performing a specified action at a specified time. Diana, and not John

and Mary, exerts a comprehensive influence in designing the zygote. It is with respect to this comprehensiveness that Diana resembles the operation of blind deterministic forces in their totality. It is for this reason step two of the Zygote Argument follows from step one. If the causal influence exerted by Diana were not comprehensive, it would not resemble the causality of blind forces in a deterministic universe. John and Mary's mental states and intentions in creating a zygote hold very limited sway over precisely what it will do many years later once it has developed. But Diana's knowledge and power *is* able to insure what the developed zygote will do at a precise time in the future. This Fischer's objection to the Zygote Argument does not succeed.

Stephen Kearns has also criticized the first premise of the zygote argument.

First, he offers the following interpretation of the premise.

1b. Because Ernie was manipulated into performing all of his actions by some agent's creating his zygote, Ernie is not a free agent and is not morally responsible for anything. (Kearns 2011, p. 5)

Based on this interpretation, he offers the following objection.

The idea, then, is that Ernie is unfree because he is manipulated since his creation. Furthermore, this manipulation is not significantly different (regarding freedom and responsibility) from causal determinism. Therefore, causal determinism is incompatible with free will. What should we make of this argument? One problem with it is that, if the conclusion is true, Ernie's lack of freedom is explained by the fact that all of his actions are deterministically caused. This being so, we may think that it is exclusively explained by this, and that it is therefore not explained by his being manipulated (thus rendering 1b false, and the argument unsound). If one is to show that Ernie's being manipulated is indeed an independent explanation of his lack of freedom, one needs a case in which Ernie is unfree because he is manipulated, but also in which Ernie's actions are not deterministically caused.

Kearns then references the previously quoted passage that explains Diana's creation of the zygote. He replies to this scenario by citing Mele's defense of premise 2 and arguing that it conflicts with premise 1.

Unfortunately, Mele's defense of premise 2 rests heavily on the fact that, in his described case, Ernie exists in a deterministic world. In effect, Mele asks us to imagine another deterministic world in which Ernie is produced normally but still does exactly the same things as he does in the original deterministic world. He then claims that this makes (or should make) no difference to our judgements about Ernie's freedom. This is partly justified by the fact that Ernie has no say over the laws of nature or properties of the universe at the time of his creation. In both deterministic universes, these combine to determine Ernie's entire future. None of this applies, however, to our new argument. In the new case, Ernie is manipulated but his actions are not determined by this manipulation. Thus even though Ernie has no say about the laws of nature, and no say about the universe at the time of his creation, these do not determine what he goes on to do. If the manipulation really explains why Ernie is not free in the new case, it does so in a substantially different way than does the obtaining of causal determinism (perhaps it does so precisely by making Ernie a tool of some other agent). Thus if we appeal to 1b (and our new case) in formulating the zygote argument, we are not justified in asserting premise 2. (Kearns 2011, pp. 7-8)²

Kearns' argument rests on his interpretation of premise 1 as saying that Ernie is manipulated in a manner that is indistinguishable from deterministic causation. This permits Kearns to argue first that Mele needs a new argument in which Ernie is manipulated and yet not causally determined. Next, with this modification of the argument, Mele is blocked from asserting premise 2 because it relies on the assumption of causal determinism in premise 1.

I concede that as formulated the Zygote Argument is subject to this criticism. But it is possible to rescue the argument by removing a single offending word. With the removal of this word, the spirit of the argument will be retained, and the objections offered by Kearns can be defeated. The word to be removed from premise 1 is 'deterministic'. Thus premise 1* should replace premise 1:

² Kearns offers several other criticisms of the zygote argument based on other interpretations of the first premise. As I regard 1b as most closely reflecting Mele's intent, I focus on this criticism exclusively.

(1*)Because of the way his zygote was produced in his universe, Ernie is not a free agent and is not morally responsible for anything.

The revised premise permits Mele to argue that the fact that Ernie develops in a deterministic universe is something one observes only *after* he is seen to be unfree. It allows him to clearly distinguish premise 2 from premise 1, for it is only at premise 2 that the resemblance between Ernie's actual development and normal development in a deterministic universe is established. What the revised first premise says is that the mere fact of Ernie being manipulated is sufficient to preclude his being morally responsible. The mere fact that a virtually omniscient goddess programs Ernie in just such a way so that at a particular time in the future he will perform a certain action suffices to show that he is not free or morally responsible. This observation is logically independent of the fact that the way he was manipulated is indistinguishable, in relevant respects, from the way a zygote normally develops in a deterministic universe. Thus with the premise 1* in place it is not necessary for Mele to suppose that the manipulation takes place in a nondeterministic universe. He merely has to assert that manipulation is present of a sufficient degree to preclude moral responsibility or freedom. He can then delay pointing out the resemblance of the actual case to how a zygote normally develops under causal determinism until premise 2. With this revision in place, the Zygote Argument is very persuasive in establishing that free will is incompatible with determinism.

Conclusion

In summary, we discussed two arguments for incompatibilism, the famous Consequence Argument and the more recent Zygote Argument. We discussed how these arguments could be defended against objections. Without claiming that these

arguments are absolutely decisive, it is nonetheless appropriate to set aside compatibilism in examining whether or not we have free will. With compatibilism out of the way we can focus on whether the prominent incompatibilist, or libertarian, theories of action are successful in securing free will. If they fail, then it may be concluded that we do not have free will.

CHAPTER 2 AGENT-CAUSATION

The first family of libertarian theories to be considered are the *agent-causal* accounts. Typically, these views require (1) that a directly free decision or action be caused by the agent, (2) that the decision (or other action) not be causally determined by any event (such as a reason), (3) that the *causing* of the decision or action also not be causally determined by any events, and (4) that the causation by the agent not consist in causation by events. Agent-causal accounts are centered upon the agent, who is considered a persistent substance; hence agent-causation is causation by a substance. Arguably, since a substance is not the kind of thing that can be an effect,¹ she can be the originator of actions, an uncaused cause. Thus agents themselves, not any events involving the agents such as reasons are the ultimate sources of decisions and actions. Recent philosophers who have propounded views of this sort are Roderick Chisholm, Timothy O'Connor and Randolph Clarke. I will summarize the respective accounts and articulate a variety of objections to them.

The Basic Account—Chisholm

The earliest agent-causal account to be examined is that of Roderick Chisholm. In the 1964 Lindley Lecture he distinguished between two different kinds of causation, causation involving events in contrast with causation by substances.

If we consider only inanimate natural objects, we may say that causation, if it occurs, is a relation between *events* or *states of affairs*. The dam's breaking was an event that was caused by a set of other events—the dam being weak, the flood being strong, and so on. But if a man is responsible for a particular deed, then, if what I have said is true, there is some event,

¹ In referring to a substance, we are speaking of a *standing entity*, an abiding existent. In itself it is not an event, and thus not an effect. Events in the life of a substance, such as its birth, can be effects, but not the substance itself.

or set of events, that is caused, *not* by other events or states of affairs, but by the agent, whatever he may be.

I shall borrow a pair of medieval terms, using them, perhaps, in a way that is slightly different from that for which they were originally intended. I shall say that when one event or state of affairs (or set of events or states of affairs) causes some other event or state of affairs, then we have an instance of *transeunt* causation. And I shall say that when an *agent*, as distinguished from an event, causes an event or state of affairs, then we have an instance of *immanent* causation. (Chisholm 1964, p. 28)

A possible difficulty for Chisholm's distinction between the causation of events (*transeunt* causation) and that by substances (*immanent* causation) is that he treats it as dividing inanimate objects from agents. But there is no reason to think that only agents can be substances. There are many kinds of substances in the world, and if agent substances are capable of being causes, then so should non-agent substances. In "the hammer flattened the tack" the fact that the hammer moved forcefully is the event that is responsible for the flattening of the tack. But insofar as the hammer is a substance, it is equally true that the hammer itself is the cause of the flattening. Perhaps Chisholm ought to distinguish between event causation and substance causation in general, which would include agents, but extend to other substances as well. As we shall see momentarily, a possible defense to this objection is contained in his characterization of the origination of action.

Chisholm argues that the agent is an originating cause.

The nature of what is intended by the expression 'immanent causation' may be illustrated by this sentence from Aristotle's *Physics*: 'Thus, a staff moves a stone, and is moved by a hand, which is moved by a man.' (VII, 5, 256a, 6-8) If the man was responsible, then we have in this illustration a number of instances of causation—most of them *transeunt* but at least one of them *immanent*. What the staff did to the stone was an instance of *transeunt* causation, and thus we may describe it as a relation between events: 'the motion of the staff caused the motion of the stone.' And similarly for what the hand did to the staff: 'the motion of the hand caused the motion of the staff'. And, as we know from physiology, there are still

other events which caused the motion of the hand. Hence we need not introduce the agent at this particular point, as Aristotle does—we *need* not, though we *may*. We *may* say that the hand was moved by the man, but we may also say that the motion of the hand was caused by the motion of certain muscles; and we may say that the motion of the muscles was caused by certain events that took place within the brain. But some event, and presumably one of those that took place within the brain, was caused by the agent and not by any other events. (Chisholm 1964, pp. 28-29)

When a staff moves a stone, it does not originate the event (the stone's moving); it is itself moved by an agency. This agency is rightly said to be the originator of the event.

At first glance this seems to uphold Chisholm's identification of substance-cause with agent-cause. But on further inspection it is clear that many sorts of substances may be said to operate as causes without any involvement by an agent. For example, a supernova can be said to cause the distribution of elemental particles, but no agent is involved. In general, many natural substances are recognized as causes, so the initial criticism holds. Fortunately, the theory can stand with 'substance' substituted for 'agent' inasmuch as all agents are substances. Therefore, in what follows we shall read Chisholm as confining his discussion to those substances which are also agents.

Before looking at additional criticisms, let us examine the remainder of Chisholm's account. He writes

Perhaps there is less need to argue that the ascription of responsibility also conflicts with an indeterministic view of action — with the view that the act, or some event that is essential to the act, is not caused at all. If the act — the firing of the shot — was not caused at all, if it was fortuitous or capricious, happening so to speak "out of the blue," then, presumably, no one — and nothing — was responsible for the act. Our conception of action, therefore, should be neither deterministic nor indeterministic. Is there any other possibility?

We must not say that every event involved in the act is caused by some other event, and we must not say that the act is something that is not caused at all. The possibility that remains, therefore, is this: We should say that at least one of the events that are involved in the act is caused, not by any other events, but by something else instead. And this

something else can only be the agent — the man. If there is an event that is caused, not by other events, but by the man, then there are some events involved in the act that are not caused by other events. But if the event in question is caused by the man, then it is caused and we are not committed to saying that there is something involved in the act that is not caused at all. (Chisholm 1964, pp. 27-28)

At the heart of Chisholm's theory is the claim that agency is rooted in causation by a substance. In the case of human beings, the substance in question is the person herself. According to this view, it is incorrect to say that actions by agents are uncaused. Rather they are caused by something that is not an event, namely the substance that is the agent. In the production of an action, many events are involved. For example, in the knocking over of a bowling pin while bowling, there is the event of the ball's striking the pin. Prior to that, there is the motion of the ball as it travels down the lane. At a still earlier stage, there is the motion of the bowler's arm. But even this is not the origin of the effect in question. We must look further back to the bowler's arm movement and beyond that to the neural event that triggers the arm movement. At this point it is difficult to identify any relevant event that is more basic to the production of action. One might point to the biochemical reactions that the neural event consists in, but as these are constitutive of the neural event, they are not really separate events but a finer grained description of what is going on at the neural level. But according to Chisholm the neural event is not the first cause in the knocking over of the bowling pin. Rather, the original cause is the agent herself. The person herself causes the neural event which leads to all the other events in the causal chain.

Objections to the Basic Account

One possible objection to the above account is put forward by Chisholm himself in a rhetorical question. He asks, "What is the difference between saying, of an event A,

that A just happened and saying that someone caused A to happen” (Chisholm 1964, p. 30)? In other words, it may be thought that in the absence of a reductive account of agent-causation the claim that a person qua substance causes an event is devoid of content. It may seem to amount to no more than saying that the event, A, just happened. However, Chisholm has a satisfactory reply. He says that in the case of agent-causation an irreducible causal relation exists between the agent and A whereas when A just happens, no such causal relation obtains. For example, if A is a certain neural event that happens by itself, then A is the beginning of the causal chain and ex hypothesi will happen by itself, uncaused. But in the case of agent-causation the beginning of the causal chain is not A, but the agent: the agent causes the neural event. Moreover, there is a real, and fundamental, causal relation between the agent on the one hand and the neural event, A, on the other.

A stronger objection to Chisholm’s account concerns the temporality of agent-causation, which I term the Broad-Ginet objection.

On the agent-causation theory, the immediate cause of the occurrence of a particular sort of simple mental event at a particular time is the agent herself, *per se* and not in virtue of any event of which she is the subject. But the agent *per se* cannot *explain* why the event happened precisely when it did rather than at some slightly different time. Only some difference between the agent at one time and the agent at the other times, some temporally located property, could do that. Nor, it might be added, can the agent *per se* explain why that particular sort of event rather than some other sort happened just then. What sense can it make, then, to say that the agent as such is the *cause* of the occurrence of that particular sort of event rather than some other sort, and is the cause of its occurring at that particular time rather than at some other time? (Ginet 1997, pp. 93-94)²

² This objection was originally formulated by C.D. Broad in “Determinism, Indeterminism, and Libertarianism,” page 169.

As a substance, the agent abides over a span of time. Yet an action occurs at one time and not another. On an explanation of action based on events, it is easy to see why an action takes place at the particular time it does. It is because it is preceded by the event that causes it. For example, we can explain the agent's causing a ball to move at time t by citing the agent-involving event at t_n of the agent's hitting it with a bat. Put another way, an action happens for a reason. But if the reason were the agent herself, then the action would be temporally coextensive with the existence of the agent. This implies that the real reason for an action at t must be something that begins to occur just before t (or possibly simultaneously with it). And this means that the substance, or agent, which is responsible for the action, cannot by itself be the cause of the action. Put another way, the agent qua substance is an undated existent. But the action occurs at a particular time. Hence the cause of the action must be time indexed, it must be dated. While the agent definitely plays a major role in the production of action, a dated agent-involving event is also required in order to explain why the action took place when it did.

This objection appears decisive against Chisholm's version of agent-causation. It remains to be seen whether more recent accounts fare better against it.

A further objection to the basic agent-causal account is put forward by Kane:

Nor does nonoccurrent causation [agent-causation] 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, such as dualist theories or explanatory incompatibilism, 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. What agent-cause theories add is that the agent (nonoccurrently) caused A at t rationally and voluntarily, and if the agent had done otherwise, the agent would have (nonoccurrently)

caused something else (or nothing at all) at t rationally and voluntarily. But this adds no further information about why the agent did it. One can understand why Hobbes thought agent-cause theories were empty" of explanatory content and why modern critics such as Gary Watson, John Bishop, and Bernard Berofsky argue that agent-causation is at best a "label" for "what libertarians want," rather than the theory they need. (Kane 1996, p. 121)

This "emptiness" objection is more general than the temporality objection. One way of putting it is that agent-causal accounts fail to fully explain why one acts in one way rather than another. At the heart of the agent-causal account is the bare nonoccurrent causing of an action. That is, the action is produced by the agent qua substance rather than via agent-involving events such as reasons. Although the agent acts voluntarily and rationally, the explanation of the action is the mere fact that the agent exists, such that if the agent were to act otherwise, or do nothing at all, the alternatives would also be rationally and voluntarily be produced, nonoccurrently, by the agent. For example, the basic agent causal account cannot answer why the agent performs an action using her left hand rather than her right, or vice versa. All that can be said is that the agent as substance produces the action. As we shall soon see, other versions of agent-causation are capable of handling this objection.

Agent Causation and Reasons Explanation—O'Connor

In *Persons and Causes*, Timothy O'Connor presents another version of agent-causation in terms of the reasons the agent acts on:

The following general conditions, I maintain, are enough to explain an action in terms of an antecedent desire. The agent 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-so-act-here-and-now-to-satisfy- Θ ;

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 causally sustained the completion of the action.

As indicated earlier, I apply the term "decision" to the core agent-causal event noted in (2), the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-A-here-and-now-to-satisfy- Θ . (O'Connor 2000, 86)

Suppose prior to swinging a bat the agent has the desire to hit a ball and believes that swinging the bat will satisfy her desire to hit the ball. Further suppose that the swinging of the bat results from her self-determination including the coming to be of the intention to hit the ball. Moreover, concurrent with the swinging of the bat the agent continued to desire to hit the ball and intended to do so by swinging the bat. Then the concurrent intention to swing the bat was a direct causal consequence, or continuation, of the agent's action-triggering intention to hit the ball, and, finally, it causally sustained the completion of the action of swinging the bat.

One question that arises in respect to the coming to be of the intention is that it seems to be an event. If so, then it may appear that this is an event-causal rather than an agent-causal account. The solution to this puzzle is that the reasons, or desires, serve only to *explain* the action in question; they do *not cause* it. The agent herself causes the action. The existence of the intention merely *explains why* the action occurred when it did. Thus O'Connor's account satisfactorily answers the emptiness objection leveled by Kane. The above considerations also suggest a possible defense against the temporality objection raised by Broad and Ginet.

O'Connor writes:

The alleged problem is that the agent causationist is committed to the possibility that a cause of an event doesn't explain the event. In reply, I begin by noting that we have to be careful here to identify just what event

is in question. Suppose, as the agency theory has it, that I cause a state of intention to call my friend here and now. Let us give the name 'e' to the event of that state of intention's coming to be at time *t*. I am at *t* the cause of *e*. Yet I also figure into the explanation of *e*, insofar as I had reasons for so acting.

"What about the event, *e's occurring at t rather than at t1*? *Ex hypothesi*, there is no explanation of this event, and yet you caused it, according to the agency theory." Here I think it is a mistake to hold that there is an event that is *e's occurring at t rather than at t-1*, in addition to *e* itself. There is no reason to hold that corresponding to every contrastive fact about a contingent occurrence there is a distinct 'contrastive event,' where an event is understood to be a concrete entity. (O'Connor 2000, pp. 75-76).

O'Connor is arguing that the temporality objection loses its force when we look closely at the individuation of events. There is an event, *e*, which is the coming to be of a certain state of intention to immediately act in particular way. There is a contrastive explanation of *e*, namely the reasons the agent has for acting in the relevant manner.³ Thus when one asks why the agent caused *e* at *t*, all one has to do is cite the agent's reasons. But the objector then challenges that there is no contrastive explanation for why *e* occurred when it did. O'Connor's reply is that "There is no reason to hold that corresponding to every contrastive fact about a contingent occurrence there is a distinct 'contrastive event,' where an event is understood to be a concrete entity" (ibid.). In other words, *e-occurring-at-t* is not a concrete entity over and beyond *e*.

In appraising O'Connor's reply we should note that there are two distinct issues here. One concerns the *rational explanation* of *e*, the other concerns the *causing* of *e*. The latter asks for a specific kind of explanation, a causal one. O'Conner may be correct in stating that *e* is *rationally explained* by the agent's reasons. But on an agent-causal account, *e* is *not caused* by the agent's reasons, but by the agent herself.

³ A contrastive explanation explains why something occurred when it did as opposed to some other time.

Therefore, O'Connor's has not yet accounted for the causing of *e* when he cites the agent's reasons. He still needs to explain why a substance caused an event at one time rather than another. The distinction between *e* and *e*-occurring-at-*t* collapses into the simple question of why *e* occurred when it did. To ask this question is not to ask about a concrete object over and above *e*, but merely to ask about *e* itself why it occurred at *t* rather than *t_n*. For example, suppose *e* is the swinging of a bat. The rational explanation for why the bat was swung is that the agent wanted to hit the ball. However, on pain of losing the agent-causal core of the theory altogether, the existence of the reason does not explain why the agent caused the swinging of the bat at precisely that the time it was swung. The agent herself, not her reasons, is the cause of the swinging of the bat. It thus remains open to ask why the bat was swung when it was. But in asking why the bat was swung when it was, we don't posit an additional entity to the swinging of the bat, we merely ask why *it, e*, happened when it did. There is merely the demand for a contrastive *causal* explanation of *e*. And from O'Connor's failure to supply this it appears clear that he fails to rebut the Broad-Ginet objection.

Further Objections

A further objection that can be applied to O'Connor's theory is Ginet's argument that agent causation leads to an infinite regress:

It is clear, by the way, why the agent-causation account does not want to allow that any extrinsic condition is needed to give the agent control over her causing *e*. For if it did allow this, then it would be in serious trouble. The only extrinsic condition the agent-causation account has to suggest—the only thing consistent with its idea as to what agent control consists in—is that we posit a further agent-causal relation, this time between the agent and the event of her causing *e*. But the same question can again be raised about the agent's causing of her causing *e*: what gives the agent control of *that*? Thus the account would either fail to give any ultimate answer to the question of what agent-control consists in or

else us the very implausible answer that it consists in an infinite regress of agent-causings. (Ginet 1997, p. 93)

This does not appear to be an effective objection. Ginet is arguing that the agent-causation theorist needs to posit an infinite number of principles of control whereby the agent produces *e*. If this were the case, the agent-causation account would be unintelligible. But there seems to be no reason why the agent herself could not be regarded as the ultimate, primal source of control over *e*. That is, there is no reason why the agent-causationist has to posit an extrinsic condition at all. In this regard it should be noted that on the agent-causal view, the *e* over which the agent exerts immediate control is the neural event that triggers the more complex action. Complex actions such as hitting a ball, or even gripping a bat, are only indirectly controlled by the agent. With this in mind it does not seem problematic that an agent could exercise immediate, direct, initiatory control over her brain states. And since there is no need to posit an intermediate control factor, there is no ensuing regress, much less an infinite one.

A more serious problem faced by agent-causation, as well as other libertarian theories, is the luck objection.

Assume that Tim chose freely in the scenario under consideration. Then, on O'Connor's view, Tim "had the power to choose to continue working or to choose to stop, where this is a power to cause either of these mental occurrences. That capacity was exercised at *t* in a particular way (in choosing to continue working), allowing us to say truthfully that Tim at time *t* causally determined his own choice to continue working." Suppose that the position reported in the preceding two sentences is true. Why should we suppose that the following cross-world difference is not a matter of chance or luck: that Tim exercised the capacity at issue at *t* in choosing to continue working rather than in choosing to do something else, as he does in some possible worlds with the same past and laws of nature? Grant that Tim "causally determined his own choice to continue working." Why aren't the differences in his causal determinings at *t* across worlds with the same past and laws of nature a matter of chance or luck? Tim was able to causally determine each of several choices, whereas a counterpart who fits the event-causal libertarian's picture was able to make – but not to

causally determine – each of several choices. If it is a matter of chance that the latter agent chooses to keep working rather than choosing to do something else, why is it not a matter of chance that the former agent causally determines the choice he causally determines rather than causally determining a choice to do something else? (Mele 2005, p. 387)

Consider Sandra. At a certain moment she agent-causes a decision to enter a room to retrieve her glasses, and until she does there remains a chance that she will instead, at that moment, agent-cause a decision to not enter the room. Thus there is a possible world that is exactly like the actual world up until the time at which Sandra agent-causes her decision to enter the room but in which, at that moment, she agent-causes a decision to not enter the room. Nothing about the world prior to the moment of the agent-causing accounts for the difference between Sandra's causing one decision and her causing the other. Ex hypothesi, in both worlds Sandra has the same reasons to enter or not to enter the room. This difference between worlds, then, is just a matter of luck. And if this difference is just a matter of luck, it does not seem to be the case Sandra is responsible for her decision. Responsibility seems to require that the agent exert control over which outcome emerges. And if responsibility is wanting the action cannot be considered free. Quite possibly, there are objective probabilities corresponding to each of the alternatives available to Sandra, but within those fixed parameters, which decision is selected is, as far as the agent's direct control goes, a matter of chance. On the agent-causation theory, therefore, the agent fails to have sufficient control over her decisions for them to count as free.

At this point the agent-causation theorist may rightly observe that the luck objection equally applies to other accounts such as event-causal libertarianism, etc. But while this reply would be valid when confronted by other versions of libertarianism, it is ineffective when libertarianism in general is being called into question. The fact that

agent-causation is not the only theory vulnerable to the luck objection cannot then count as a satisfactory defense of it.

A Hybrid Account—Clarke

Clarke's theory is sometimes classified among agent-causation accounts, but is actually a hybrid that employs both agent-causal and event-causal elements.

The suggested account may be illustrated as follows. Suppose that on some occasion a certain agent, Diana, is deliberating about whether to pursue a certain course of action *A1* or an alternative *A2*. She has reasons favoring each and an intention to make up her mind now. Suppose that there is a nonzero probability that *R1*—Diana's having the reasons favoring *A1*—(together with her having the indicated intention) will nondeterministically cause, in an appropriate way, her making a decision *D1* to pursue *A1*; and suppose that there is, as well, a nonzero probability that *R2*—Diana's having the reasons favoring *A2*—(together with her having the intention) will instead nondeterministically cause, in an appropriate way, her making a decision *D2* to pursue *A2*. Then, given all prior conditions, it is genuinely open to Diana to make the former decision and genuinely open to her to make the latter instead. Now suppose that, as a matter of nomological necessity, in the circumstances, whichever of the open decisions the agent makes, that decision will be made, and it will be caused by the agent's having the reasons that favor it (together with her having the intention to make up her mind), only if the agent causes that decision...Finally, suppose that, in fact, Diana makes decision *D1*. Her decision is caused by her, and it is nondeterministically caused, in an appropriate way, by *R1* and her having the present-directed intention to make up her mind. On the proposed integrated agent-causal view, the agent's exercise of direct active control consists in her action's being caused, in this way, by her and by these sorts of agent-involving events. (Clarke 2003, p. 136)

Clarke's strategy here is to endorse an event-causal theory of action in general. The agent-causal component is what accounts for *free* action. A crucial element of this account is that an agent's decision is free only if it is *nomologically* necessary in the sense that the decision she makes will be caused by her having the reasons for which she acts if and only if it is caused by her. The reason for invoking nomological necessity rather than conceptual necessity is to keep the account from collapsing into an event-

causal account. Conceptual necessity here would mean that the free action would consist simply of the action being done for reasons, which would be an event-causal view. At the same time, the view has to require at least nomological necessity or else it would be merely a contingent fact that agent-causation and causation by reasons co-occur in free actions. But if that were the case, then either what the agent's reasons cause would not be a matter of what she agent-causes, or what the agent causes would not be a matter of what she has reasons to do. And this in turn would render the agent causally redundant. These considerations show that it is nomological necessity which integrates the agent-causal and the event-causal dimensions of the theory. As a result, the agent's reasons and the agent's causality work in lock-step with one another.

Objections to Clarke's Account

Let us reflect on how well Clarke's approach fares against the Broad-Ginet temporality objection. According to agent-causation theory, the agent herself and not any agent-involving events such as reasons are responsible for the production of actions, whether overt actions or mental actions such as decisions. But the agent is an abiding substance. The mere existence of this substance cannot explain why an action occurs when it does. Instead, some dated difference within the substance seems required to explain why an action occurs at a particular time. In general, agent-causation cannot accommodate such a dated difference since it makes the agent qua substance the total cause of the action. On Clarke's account, however, the total cause of the action is not limited to the agent herself. Rather, the total cause is comprised of both dated and undated elements. The undated element is the agent; but there are also reasons R1 and R2 which are dated. Thus it is possible that the hybrid character of the theory immunizes it against the temporality objection.

Whether this reply is successful depends on how seriously we take the role of the non-dated component of the account, viz., the agent herself. A reasonable interpretation of the nomological necessity requirement is that the decision the agent makes will be caused by her having the reasons for which she acts *if and only if* it is caused by her. This suggests that the causality of the agent is the decisive causal factor. Admittedly, the dated component, reasons R1 and R2 are necessary for the production of action, but they are not sufficient by themselves. Rather, they are contributing causal factors. Ultimately, the agent herself must *cause* the action in the sense of being the decisive ingredient of the total cause responsible for triggering the action, or else Clarke's account would be indistinguishable from event-causal libertarianism. Unless the agent exerts distinct causality, we don't have an integrated account at all. But with this in mind now we can ask why the action occurs when it does. The reasons are admittedly necessary factors. The mere presence of the reasons will not bring about the action. Nevertheless, the presence of the reasons do explain why the action occurs when it does. The temporality objection is thus ineffective against Clarke's hybrid account.

A specific objection to Clarke's account is supplied by Hilary Bok. She writes:

We understand what it means to be someone's sister. But it does not follow that we understand what it would mean to be the sister of an event. It would not help to be told that our relation to such an event would be the exact same relation that we now stand in to our siblings. What we need, rather, is an explanation of how we could stand in that relation to an event: how, for instance, our mothers might in some nonmetaphorical sense have given birth to one. Likewise, we cannot assume that it makes sense to say that agents can stand in the same causal relation to events that other events do, absent some explanation of how an agent can produce an event in some way that is not reducible to event causation. (Bok 1998, pp. 44-45)

This objection draws an analogy between the sibling-sibling relationship and the relationship between an agent and an event. The relationship between one event and

another is well understood, as is the relationship someone has with her sister. But just as a sibling relation cannot, without further explanation, be intelligibly applied to an event, to say that an agent, unmediated by events, can be the author of an event, stands in need of explication. In other words, it is not immediately clear how an agent could be the cause of an event without there being any events involved in the causation. For example, in the case of an agent hitting a home run there are many agent involving events, including the swinging of the bat, or causing the bat to collide with the ball, and so forth. It is more difficult to see how the agent could, tout court, be the cause of the home run.

In Clarke's defense, Pereboom suggests that it is not un-intuitive that choices can be caused by agents rather than by events. Moreover, in the case of certain choices such as my choosing chocolate over strawberry ice cream, causation by an agent seems more natural than causation by an event (Pereboom 2001, p. 61).

While it is true that agents are often cited as the causes of this or that, it is arguably the case that agent-involving events are always responsible, upon analysis, for the causation. In the case of choosing chocolate over strawberry, a mental state of the agent, such as a desire, is normally responsible for the decision. Even in a case where the agent does not prefer one over the other, the arbitrary firing of one set of neurons rather than another would be the event that most plausibly is responsible for the decision. The reason agents or substances in general, are never causes in and of themselves is because a substance is an abiding fact, an undated continuity. Insofar as it simply exists, there is nothing about it that can cause anything. In order for causation to take place, some difference in the substance must arise. This difference is what we

term an event. For this reason, whenever an agent or substance is said to be the cause of something, it is an elliptical statement that omits the difference in the substance (i.e., the substance-involving event) that is the immediate cause. Agents are certainly causes, but indirectly via events. Thus Bok is correct in pointing out the puzzling character of treating agents as the immediate causes of events. As she concludes, “we cannot assume that it makes sense to say that agents can stand in the same causal relation to events that other events do, absent some explanation of how an agent can produce an event in some way that is not reducible to event causation” (Bok 1998, p. 45).

Metaphysical Considerations—O’Connor

From the earliest version of agent-causation theory, a question has lurked in the background: What is the ontological status of the agent? As we have seen, agency does not coincide with the ordinary embodied person. Agency cannot coincide with the ordinary person because the agent is said to be able to directly cause her brain states. Consequently, it is natural to infer that agent-causation presupposes *substance dualism*, whereby a nonphysical self is the initiator of all actions attributable to the agent. Obviously, this is an unattractive option because of the generally accepted implausibility of this version of dualism. Hence it is to the advantage of agent-causation that O’Connor eschews substance dualism in favor of *property dualism* as the source of agency. More precisely, he holds that the causal power of the agent is an emergent property distinct from the natural properties of the agent, such as the agent’s neural properties.

I have just argued that the emergence of phenomenal consciousness is a good bet. The agency theorist is committed (given the substance monism that the Causal Unity Thesis strongly suggests) to the emergence of a

very different sort of property altogether. Instead of producing certain effects in the appropriate circumstances itself, of necessity, this property enables the individual that has it in a certain range of circumstances to freely and directly bring about (or not bring about) any of a range of effects. This further commitment leaves the theory's proponent open to a special objection, not applicable to emergentist claims generally: given the unique nature of the type of property the theory postulates, it is doubtful whether it could emerge from other natural properties. It will be claimed that this property would require a very different kind of *substance* than material substances, as is posited by Cartesian dualism. (It is noteworthy that many philosophers who discuss the agency theory seem to simply assume that its adherents are dualists.)

This argument also does not bear well under scrutiny. Given that there is nothing inconsistent about the emergence of an "ordinary" causal property, able to causally influence the environments in which it is instantiated, it is hard to see just why there could not be a variety of emergent property whose novelty consists in enabling its possessor to directly effect changes at will (within a narrowly limited range and in appropriate circumstances). (O'Connor 2000, p. 121)

As I understand him, for certain individuals such as human beings, their natural properties are so constituted that under suitable conditions various higher order properties emerge. These properties are distinct from the natural properties they emerge from. On O'Connor's account, both phenomenal consciousness and an initiatory causal power are examples of such emergent properties. One advantage of this theory is that there is no need to posit a problematic and controversial substance to account for the causal power. Rather, at a suitable level of complexity, matter itself acquires new properties. And among them is the power of causal initiation.

Objections

Let us take a closer look at the emergentism O'Connor proposes. Jaegwon Kim has articulated *the causal inheritance principle* which spells out the relation between higher and lower level properties:

If mental property M is realized in a system at *t* in virtue of physical realization base P, the causal powers of *this instance of M* are identical with the causal powers of P. (Kim 1993, p.326)

According to this principle, any higher-level causal power at a given time will be token identical to a lower-level causal power. This follows in virtue of it being the case that (a) there is only one kind of substance under consideration, and (b) the higher causal power' inability to magically "free-float" and instead having to be tied to a supervenience base. The nomic connections at the higher level are therefore traceable to lower level nomic connections.

Next, let us consider what explains what. In the case under consideration we want to know what explains the causality of the agent. Construed as an emergent power, does agent-causation provide a sufficient explanation for the relevant effects, or does it depend on something else?

Suppose, for some given event, you have an explanation and I have another, distinct explanation. It can be rational, from an epistemological point of view, for you to accept yours, and for me to accept mine. But can they both be "correct" or "true" explanations? The metaphysical principle of explanatory exclusion says this: they can both be correct explanations only if either at least one the two is incomplete or one is dependent on the other. (Kim 1989, p. 95)

As expressed here, the principle of explanatory exclusion requires that we view the causal power of the agent as dependent on the ordinary, natural properties of the agent. Thus insofar as we can explain an action by saying the agent caused it, this explanation depends in turn on a more basic explanation in terms of the natural properties of the agent qua ordinary individual. The upshot of these considerations is that an emergent power of an agent's natural powers cannot be the complete explanation for an agent's actions. The emergent powers can only be intermediate or partial explanations. Moreover, this also means that the alleged emergent powers cannot explain how an agent causes her own neural states. For the emergent causal powers of the agent must be dependent upon the causal powers of the underlying physical states. And this in turn

makes the causal powers of the agent dependent on ordinary event causation.

Causation by the agent, over and above causation by events, disappears.

In spite of the above considerations, it remains open to the agent-causationist to “bite the bullet” and reject either the causal inheritance principle or physicalism. If she takes either of these options, she faces a strong empirical objection.

At this stage, then, two proposals for the nature of agent-causal powers remain, strong emergentism and non-physicalism. Against each of these views, we might raise the second sort of empirical challenge we discussed earlier: Given our scientific understanding of the world, how could there be agent-caused decisions that are freely willed in the sense required for moral responsibility? It would seem that on either of these views, if agent-causes are to be capable of such free decisions, they would require the power to produce deviations from the physical laws – deviations from what these laws would predict and from what we would expect given these laws. But such agent-causes would be embedded in a world that, by the evidence that supports our current theories in physics, is nevertheless wholly governed by the laws of physics. According to this second type of empirical objection, then, the claim that there are agent-causes is not credible, given this evidence. (Pereboom 2001, p. 79)

If the agent-causation theorist insists that the agent’s causality is truly independent of natural properties either by rejecting causal inheritance and embracing a strong version of emergentism, or by accepting substance dualism, she would be ascribing to agents the power to deviate from established physical laws. Specifically, such positions would deny physical causal closure. That is, they would deny the principle that physical events can be given complete causal explanations that refer only to other physical events. In addition, the agent-causationist would be claiming that a causal factor not accounted for by physical law would be able to intervene in physical reality in order to produce actions. Presumably, the actions immediately produced by such extra-physical interventions would be changes in the neural states of the agent. These interventions would produce

changes in the neural system that could not be accounted for by the preceding physical state of the nervous system.

It should be admitted that there is nothing logically incoherent about such a position. Nevertheless, both strong emergentism and non-physicalism make predictions which conflict with what is empirically observed. We do not observe deviations in physical systems, including the nervous system, which cannot be accounted for from within the physical system(s) itself. What I have in mind are the physical conditions, including the neural substrate, that causally produce any given action. The physical causes are seen to be sufficient for the subsequent action. Thus, there are no deviations from physical causal closure. Moreover, we lack an empirical theory that could explain how such deviations would be possible. Finally, the empirical theories we do have are very effective in explaining the way the world works. Hence agent-causation turns out to be dubious on empirical grounds. As it is a desideratum for most philosophers that their theories be consistent with the best available science, we have strong meta-philosophical reasons for rejecting agent-causation.⁴

Conclusion

We have examined several version of agent-causal libertarianism in the light of several objections. We found that it was successful in resisting the “just happened” and the “infinite regress” objections. However, most were found to be vulnerable to the Broad-Ginet temporality objection, and others to the “emptiness”, “sibling”, “causal dependence”, and “empirical plausibility” objections. On balance we have strong reasons for rejecting free will of the sort described by agent-causation theories.

⁴ Arguably, the objection based on physical closure would be effective against any agent-causationist account, including Clarke’s hybrid theory.

CHAPTER 3 NONCAUSALIST LIBERTARIANISM

Some incompatibilists propose a theory of action on which free actions are neither caused by anything nor possess any internal causal structure. Because they impose no positive causal requirements on free action, they are labeled *noncausal* accounts. This distinguishes them from causal accounts such as those that take reasons to be the causes of actions. By contrast, these theories typically posit a basic mental action as the beginning of each intentional action. The most common example of a basic mental action is a choice or a decision. These in turn are constituents of nonbasic actions such as the raising of one's arm. In such a case, the nonbasic, complex action is constituted by a basic mental action bringing about a certain bodily movement. The basic action in such an instance is also termed a volition, which is the agent's trying, endeavoring or willing to move her body (or part of it) in a certain way.

A further feature typifying noncausal libertarian theories is their ascription of intrinsic activeness and purposiveness to basic actions. On such views, when one makes a decision, the intending to make it is intrinsic to it, such that when one decides to *A*, one intends to decide to *A*. Also, as Ginet describes it, basic actions such as decisions have an "actish phenomenal quality" (1990, p. 13) wherein it immediately seems to the agent *as if* she is directly producing, making happen, or determining the event that has this quality.

The main proponents of noncausal libertarianism are Carl Ginet, Hugh McCann, Stewart Goetz, and Storrs McCall. In general, these philosophers argue against causalism, and argue for particular accounts of the metaphysics, phenomenology, and

structure of action. In what follows I will examine typical noncausal approaches to each of these issues, and show why each falls short.

The Rejection of Causalism

To many philosophers, it seems intuitive that events such as reasons-states contribute crucially to the production of actions. For this reason, noncausalist libertarians attempt to show that this intuition is incorrect. To this end, Storrs McCall argues that causalist accounts of action flounder in the face of *akrasia*, weakness of will.¹

The problem for Aristotle, and for Davidson, is: how can someone know perfectly well, all things considered, that it is better to do *A* rather than *B*, and yet do *B*? This is the problem of *akrasia*, weakness of will. For Davidson, who sees the reasons as causes, there can be no solution to the problem. The incontinent man has every reason to behave continently, yet somehow these reasons do not constitute a cause. Why, Davidson asks, would anyone ever perform an action when he thought that, everything considered, another action would be better? (1980:42) And yet, of course, the *akrates* does just that. There is no *causal* explanation of why the incontinent man does what he does. (McCall 1994, pp. 270-271)

Having tagged causalism as failing to account for weakness of will, he goes on to explain how noncausalism would handle it.

The causal theory, then, cannot cope with weakness of will. But a theory of action which takes choice or decision as its basic concept has a chance of succeeding where the causal theory fails. Either (i) weakness of will takes the form of an agent's performing an action at the same time that he believes that, all things considered, it would be better to do something else, in which case the action can be accounted for...by the lack of any necessary connection between evaluation and decision, and by the primacy of decision. Or (ii) weakness of will takes the form of indecisiveness, vacillation, and lack of commitment, in which case it constitutes one of the infirmities to which decision is subject. Either way, *akrasia* is explicable, whereas it does not appear to be explicable if we adopt the causal theory. (McCall 1994, p. 271)

¹ It should be noted that the interpretation presented in the following quote would in all likelihood be disputed by Davidson.

Weakness of will is widely discussed in moral thought and the philosophy of action. If a theory is unable to effectively account for it, that is a strong reason for rejecting it.

Objections

Contrary to the above arguments it can be shown not only that a causalist account of action can explain *akrasia* but that it can utilize very similar explanations as those the noncausalist would invoke.² One line of causalist argument relies on the claim that the mind is not transparent. By this I mean that not all mental contents are accessible to introspection at all times. Equivalently, I am claiming that some mental contents are sometimes inaccessible to introspection. This fact was originally uncovered by psychoanalytic theory, but I am taking it to be an accepted empirical fact about the mind independently of the truth of psychoanalysis in general. This *non-transparency thesis* permits the causalist to say that when it appears that an agent acts contrary to her own best reasons, the agent is merely acting against what she *thinks* are her own best reasons, that her truly best reasons, the reasons she actually acts upon, are not immediately accessible to introspection. Admittedly, this reply may make it appear that seemingly rational agents are not truly so insofar as they act for reasons not sanctioned by their conscious deliberation. But this result is consistent with general observations of human behavior. Arguably, human beings are quite often less than fully rational in their deliberations and behavior.

This strategy may be opposed, however, on the grounds that it *explains away* rather than *explains* weakness of will. Applying the non-transparency thesis in all instances when an agent appears to act contrary to her best reasons would entail that

² In defending causalism I am not attempting to directly defend Davidson's theory or to suppose that Davidson would necessarily endorse what I am saying.

true *akrasia* does not exist. The strong intuition that *akrasia* is a real possibility warrants a causalist explanation of how it is that someone *in reality* acts contrary to her *best* reasons. Given this requirement, the causalist can argue instead that while reason-states are the causes of action, our introspective tracking of bestness does not necessarily correspond to the causal efficacy of the respective reasons. This is a different version of transparency. Let us compare the two theses. The one initially proposed is:

(a) *non-transparency with respect to the bestness of reasons* (NTBR) = that introspection sometimes fails to identify which of one's reasons one most strongly prefers.

In contrast, the second non-transparency thesis is:

(b) *non-transparency with respect to the causal-efficacy of reasons* (NTCER) = that introspection sometimes fails to identify which of one's reasons is causally stronger.

NTCER holds that we may sometimes decide which reason has the most merit and consciously commit to acting in accordance with it, even though some other reason which we don't commit to is causally stronger to the extent that it overrides our conscious commitment. Thus reasons are what cause actions, yet the reason which is evaluated as the best may not be the causally strongest reason. Like NTBR, NTCER implies that human beings are not fully rational in the sense that conscious deliberation and the evaluations that result from it do not always govern our conduct. But this consequence should not be troubling. Not only are human beings not perfectly rational, but *akrasia* itself is a species of irrationality. The causalist can explain this irrationality on the grounds that reasons that are valued the most are not always the causally strongest, and that this disparity is not always accessible to introspection.

By invoking NTCER, the causalist can make use of similar explanations as those suggested by the noncausalist. The first explanation offered by the noncausalist denies “the lack of any necessary connection between evaluation and decision” and prefers “the primacy of decision”. The causalist can replace this with the denial of a necessary connection between evaluation and the causal force of a reason, while retaining the primacy of reasons. The second noncausalist explanation invokes “indecisiveness, vacillation, and lack of commitment.” This option is equally acceptable to the causalist. Even though reason-states are the causes of action, an agent may be indecisive and vacillate with respect of her reasons such that no clear commitment is established. In such cases, there would be no exclusive preference for one reason over the others, a scenario perfectly compatible with reasons being the cause of actions.

In the light of the above, there is no reason to prefer a noncausalist account of action to a causalist account.

Noncausalist Phenomenology of Action

As mentioned above, noncausalists posit a certain phenomenology of action. For example, both Carl Ginet and Hugh McCann argue that free actions have a distinctive subjective character. Ginet says:

Every action, according to me, either is or begins with a causally simple mental action, that is, a mental event that does not consist of one mental event causing others. A simple mental event is an action if and only if it has a certain intrinsic phenomenal quality, which I've dubbed the "actish" quality and tried to describe by using agent-causation talk radically qualified by "as if": the 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!"). 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). (Ginet 1997, p. 89)

Ginet is, of course, correct in claiming that those actions we perform on purpose, including mental actions, have a characteristic feel or phenomenology. When we act, it seems to us that *we are doing* something, as opposed to something merely happening to us. A similar observation is made by McCann with respect to decision making:

The key to realize is that the phenomenon of deciding is *intrinsically intentional*. That is, it is not even possible for a decision to be made inadvertently or without our meaning to make it. Along with volition, decision making is by its very nature an exercise of agency. This cannot be a matter of some relation obtaining, such as that acts of decision are caused by reason-states. On the contrary: if that were so, it *would* be possible for decisions to occur without our meaning to make them, by a deviant causal chain or the like. Nor is it a matter of it entering into the content of my decision that it will be made for certain goals. We have seen that that does no good, and in fact is not even relevant. The intentionality of acts of decision is not an element of their conceptual content; it is a nonrelational and essential feature of the act of deciding itself. (McCann 1998, p. 163)

Some doubt may be raised about these claims in light of the possibility of non-conscious or sub-conscious mental processes. But he is correct that there is a class of mental actions, namely conscious decisions, which are done on purpose. Let us also leave aside for the moment the possibility of decisions being brought about by deviant causal chains. Certainly, in the normal case, decisions are intentional. Next, McCann explicitly connects the intrinsic intentional character of action to its ontological status as an action.

One has to do with ontological foundations. An exercise of agency has to be spontaneous and active; it is a creative undertaking on the agent's part, to be accounted for in terms of its intrinsic features, not via the operations of other denizens of the world. Second, exercises of agency must be intentional; they have to be undertaken for the sake of some objective the agent deems worthy of attainment. (McCann 1998, p. 180)

He argues that acts of agency such as decisions and volitions must be accounted for in terms of their intrinsic characteristics as opposed to “other denizens of the world” like reason-states.

Objections

Timothy O'Connor objects to the phenomenological justification for noncausalist libertarianism on the grounds that introspective judgments are defeasible.

Perhaps some such introspectible quality is part of our basis for distinguishing our own actions from events that merely happen to us (such as involuntary blinkings). In part because it appears to me that I have (in some sense) directly brought about this movement, I believe that I have done so. Furthermore, it is quite consistent with an event's having this phenomenal quality that it be causally undetermined (or even uncaused). Our introspective judgments that we are responsible for a given action is defeasible, however. They presuppose that things pretty much are as they seem with respect to our own behavior, and hence that its seeming as if I directly bring about some intentional state is owing to the fact that I do so bring it about. If I'm given strong reasons to believe that this is not actually the case, I would withdraw the judgment that I determined it. Hence, pace Ginet, the causal ancestry of such events is relevant to the matter under consideration. (O'Connor 1991, p.26)

Consider a scenario in which an agent's brain is tampered with so that it appears to her that she is acting freely, when, in reality, her actions are the result of external manipulation. Although this is not how actions are normally brought about, it is compatible with the noncausalist phenomenological description. This possible counterexample is sufficient to show that no phenomenal quality of an action can preclude its being brought about in an unfree manner. Contrary to McCann, in order to explain the beating of hearts and the formation of stars, etc., we must look to causes of the respective events, not their intrinsic features. In general, no intrinsic characteristic of an action is capable of establishing whether or not an action is free. Therefore it is always necessary to look to its causal basis of in order to explain it.

In Ginet's defense, his citation of the actish phenomenal quality is intended merely as an "as if" indicator of free action. In correspondence with Clarke, Ginet concedes that the phenomenological description of free action is merely metaphorical (Clarke 2003, p. 20). Also, on page 9 of his *On Action* (1990) he specifically discusses an example of an event with the phenomenal quality typically associated with voluntary actions being brought about by external stimulation. At best, then, the intrinsic phenomenal character of an action is a defeasible sign of its being free that holds under normal circumstances.

Noncausalist Metaphysics of Action

Since it is intuitive to many thinkers that actions are brought about causally through events such as reason-states, it behooves the noncausalist to explain how it is possible, metaphysically, for choices to come about without being caused to do so. That is, a noncausalist libertarian theory ought to explain what kind of being a choice is that it can emerge uncaused. A metaphysical account of free action is offered by Stewart Goetz

How can a choice be intrinsically active? In the following way: (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. Consider (i). A mental power is an ontologically irreducible property which is exhibited by an entity. Corresponding to a mental power is the exercising of that power. Exercising a mental power is acting. One kind of mental power an agent has is the mental power to choose. When an agent exercises, he chooses. Another mental power an agent has is the mental power to reason or think about propositions and their logical relations. When an agent exercises this power, he actively directs his attention to propositions and seeks to become aware of their implications in logical space. (Goetz 1997, pp. 196-197)

The key expression in the above is "ontologically irreducible property" as applied to mental powers. This phrase has dualistic implications. In order for mental powers to be

ontologically irreducible, they must stand in contrast some other category of entity.³ Presumably, that entity which mental powers are being distinguished from is matter. Thus Goetz appears to be maintaining that mental powers, of which choice is an instantiation, are ontologically distinct from matter. Moreover, in light of the prima facie relation between the mind and the brain, he is probably best read as saying that mental powers are distinct from the brain. And if this is the case, it is possible for mental powers to remain uncaused while brain processes are caused.

Objections

Strong reasons for rejecting Goetz's claim that mental powers are ontologically irreducible, uncaused events are the assumption of the supervenience of the mental on the physical together with physical causal closure. The latter is the empirically confirmed principle that physical events can be given complete causal explanations that refer only to other physical events. The former is a metaphysical premise. This premise does not beg the question against Goetz insofar as it is compatible with mental powers as being ontologically distinct from the physical, including the brain, yet it is central to the scientific world-view which most contemporary philosophers are committed to. One may consistently hold that mental powers (or properties) fall into an irreducible ontological category and that the mental supervenes on the physical, especially neural states. But to deny the supervenience of the mental on the physical, one must either deny the evidence of physiological psychology or adopt an extravagant metaphysics.⁴ Activation of specific neural circuits is followed by specific, predictable sensory functions. Likewise,

³ I am assuming that by singling out mental powers as ontologically distinct, Goetz is thereby not invoking idealism. Idealism would not need to contrast mind with any other kind of entity as it maintains that all being is mental in character.

⁴ The parallelism of the mental and the physical is an example of an extravagant view.

damage to neural circuits results in predictable deficits in cognitive and perceptual function.

In *Mind in a Physical World* (1998), Jaegwon Kim presents a powerful argument, the *supervenience argument*, which can be understood as showing that mental powers cannot operate independently of their physical base.⁵ From this it is a short step to infer that if the physical base is caused so are the mental powers that supervene on it. In other words, if the mental supervenes on the physical, then mental states inherit the causality of their physical base. Let us begin with the definition of mental-physical supervenience:

Mental properties supervene on physical properties in the sense that if something instantiates any mental property M at t, there is a physical base property P such that the thing has P at t, and necessarily anything with P at a time has M at that time (Kim 1998, p. 39).

Next, consider the fact that mental powers have to figure in causal chains whether a mental action gives rise to another mental state or whether it produces a physical action. Accordingly, suppose that an instance of a mental property, M, causes another mental property, M*, to be instantiated. Kim conjectures that there are two alternative explanations for this causal sequence. One possibility is that an instance of the first mental property causes the second directly. The other possibility is that “M* is instantiated on this occasion...because P*, the physical supervenience base of M* is instantiated on this occasion” (Kim 1998, p. 42). Kim next goes on to develop the implications of preferring the second alternative to the first.

⁵ The combination of mind-body supervenience and physical causal closure may appear to entail the epiphenomenality of mind. In the present discussion, I take no stand on this matter.

Before considering the remainder of the supervenience argument, however, it is worthwhile to consider why the first alternative, the supposition that "an M-instance caused an M*-instance," is problematic. According to the definition of mind-body supervenience, in order for M* to be instantiated at t a physical base property of M* would be instantiated at t.⁶ Moreover, the notion of supervenience (in general) is supposed to be one of the dependence of the mental upon the physical in the sense that an instantiation of a mental property requires the occurrence or instantiation of its respective base property. Thus on the assumption of mind-body supervenience it appears impossible that something could cause M* to be instantiated without bringing about the physical base property upon which it depends.⁷

Returning to the supervenience argument, Kim offers an explanation for the second alternative mentioned above: "M caused M* by causing P*. That is how this instance of M caused M* to be instantiated on this occasion" (ibid.). Since P* is the physical base of M*, in order to bring about M*, M must bring about P*.⁸ To justify this conclusion Kim proposes the general principle that "to cause a supervenient property to be instantiated, you must cause its base property (or one of its base properties) to be instantiated" (ibid.). Accordingly, in order to alleviate a headache one must intervene among the brain processes upon which the headache supervenes. Likewise, on the

⁶ For simplicity I am overlooking the possibility that M is multiply realizable. In such a case there would be a disjunction of properties, $P \vee \dots \vee P_n$, such that if something had any of these at t it would necessarily have M at t also.

⁷ Admittedly, the apparent impossibility of an instance of a mental property directly causing another mental property to be instantiated can be avoided on theories of the relationship between the mind and the brain that do not presuppose mind-body supervenience. Mind-body *parallelism*, for example, is a view that avoids this difficulty. However, I am assuming that such an option would be viewed as unattractive.

⁸ As mentioned above, we are leaving aside the multiple realizability of M*.

assumption that the beauty of a painting is a property that supervenes upon the paint and its arrangement upon the canvas, in order to make a painting more beautiful it will be necessary to causally interact with the paint. It will not be possible to make the painting more beautiful in a “more direct” manner.

Kim next observes that: “M itself has a physical supervenience base P” (Kim 1998, p. 43). Kim argues that this fact motivates taking P as pre-empting the claim of M as cause of P*. Let's say causation is understood in terms of nomological sufficiency, then since P is sufficient for M and M is sufficient for P*, P is sufficient for P*. That is, once you have P, you don't need anything else to get P*. But if P is a cause of P*, then we at least have prima facie reasons for doubting that M is the cause of P*.

Alternatively, if causation is understood in terms of counterfactuality, it still looks as though P is a cause of P*. If P had not occurred, then M would not have occurred. And if M had not occurred, then P* would not have occurred. It is reasonable, then, to conclude that P* would not have occurred if P had not. P thus seems to be the cause of P* on both a counterfactual and a nomological sufficiency understanding of causation.

Goetz, of course, holds that mental powers are *uncaused*. But we see from the preceding discussion that mental properties (or powers) are dependent on their physical base. This implies that they cannot be uncaused if their physical base is caused. Goetz might reply that the uncaused status of the mental powers is an emergent property. But this is not intelligible. A mental power cannot exist unless its base property exists. The specific physical properties of the brain form the base of the relevant mental properties. But it seems intuitive that in order for a physical property to exist, it must be caused.⁹

⁹ Even quantum events are spoken of as non-deterministically caused as opposed to being uncaused altogether.

Hence, there is no scope for the uncaused appearance of mental properties. They have to supervene on the physical effects of some physical cause(s), which entails that they figure in causal chains as either causes or effects. A further consideration is that positing emergent uncaused properties implies the violation of physical causal closure, which is not empirically observed. Yet it would have to be observed quite often given the ubiquity of mental actions. Thus the supervenience of the mental on the physical together with physical causal closure leads us to conclude that noncausalism is false.

Noncausalist Explanation of Action

Apart from metaphysical considerations, noncausalist libertarianism has to provide an explanation of action. Ginet commences his account of how reasons explain actions with the following schema.

(1) S A-ed in order to B.

(2) S A-ed because she had promised to B and she believed that by A-ing she would B. (Ginet 2002, p. 388)

He then offers two conditions which are supposed to be sufficient for the truth of (1) and (2) respectively.

(1-C) Concurrently with her A-ing S intended of that A-ing that by it (and in virtue of its being an A-ing) she would B (or would contribute to her B-ing).

(2-C) Before her A-ing, S had promised to B, and concurrently with her A-ing S intended of that A-ing that by it she would keep that promise. (ibid.)

According to the above, an action can be sufficiently explained by reference to the intention the agent has when performing it. It is thus unnecessary to cite a reason-state as the cause of the action. Suppose Jane phones John because she previously made a promise to do so. On a standard desire-belief model of action one would say that Jane's phoning John was caused by her desire to keep her promise plus the belief that phoning

(at that time) would satisfy the promise. Ginet's approach dispenses with the invoking of a cause. Instead, it would explain Jane's action by saying that concurrently with making the phone call she on the one hand intended to call John, and on the other that she intended to fulfill the promise she had made to call him. The explanation for the action is thus internal to the intentionality of the action and does not require any event, such as a reason-state, to contribute causally to the production of the action.

Objections and Discussion

In his objection to the foregoing account, Alfred Mele challenges the notion that conditions (1-C) and (2-C) are sufficient to explain action. In this example, an agent, S, has two concurrent intentions when performing the action of opening a window.

[S]uppose that a mad scientist, without altering the neural realization of N itself, renders that realization incapable of having any effect on S's bodily movements ... while allowing the neural realization of O to figure normally in the production of movements involved in S's opening the window. Here, it seems clear, O helps to explain S's opening the window, and N does not. Indeed, N seems entirely irrelevant to the performance of that action. And if that is right, Ginet is wrong; for on his view, the mere presence in the agent of an intention about her [action] is sufficient for that intention's being explanatory of her action. (Mele 1992, p. 253)

In this passage 'N' refers to the intention to let in some fresh air, and 'O' refers to the intention to gain a better view. Thus Mele says that in a case where the intention to let in fresh air is rendered causally inert, we would have to explain the action by citing the other intention on the grounds that it would be causally efficacious. This objection aims to refute Ginet's account because the latter does not say the intention one has in mind as one performs an action is causally sufficient for the action. Certainly, in normal cases it would be. But as a general theory of action it should be able to handle deviant cases. Hence, it appears to be a flaw in the account that it can't explain what produces actions when an intention is causally inert.

Ginet's responds to Mele's objection by suggesting that it is question-begging. He says while the objection presupposes that the neural realizations of our intentions causally bring about our voluntary exertions, we do not actually know this to be the case. He says that in the absence of such knowledge, we are nevertheless in a position to give noncausalist explanations of actions (Ginet 2002, pp. 389-390).

With regard to the charge of question-begging, Mele could reply that the intuitively plausible presumption of the supervenience of mental states on physical or neural states is a *prima facie* reason to suppose that the neural realizations of our intentions do indeed causally bring about our actions. The plausibility of mind-body supervenience is, of course, strongly buttressed by evidence from neuroscience and neurology. Moreover, in light of the previous discussion of the metaphysics of action, we see that mental states are not causally independent of their physical base. Hence it is to be expected that our intentions inherit the causality of the neural states they supervene on.

Although the above argument will likely convince many, it still may be insisted that even the presumption of mind-body supervenience is unproven. Rejecting it, to be sure, is costly in that it seems to set aside the scientific worldview. But it is still possible to reject it. It turns out, however, that Mele's objection can be reformulated so that it does not beg any questions about the relation of the mind and the brain. Instead of supposing that the neural realization of intention N is rendered causally inert, let us stipulate merely that N is, for some reason, rendered causally inert, period. In other words, without speculating on the reason why intention N is causally inert in a given scenario, just suppose that it is. Then as before, suppose that there is a concurrent

intention O. It follows that S will open the window on the basis of O rather than N precisely because the former would be causally efficacious. It might be objected that we don't know that O is causally efficacious, but this seems like a justified inference to make in explaining why S acts on O rather than N. Inasmuch as the two intentions are concurrent, the relevant difference between them seems to be that one of them is not causally efficacious. This shows that it is not the case, as Ginet holds, that the mere presence in the agent of an intention about her action is sufficient to explain the action. And from this it may be inferred that what is missing is a reference to causality. Specifically, what is needed is to require that the intention an agent acts upon in performing an action is also causally efficacious.

A further objection to Ginet's account refers to a scenario in which Sarah has two desires prior to entering a room (Ginet 1990, p. 145). One desire is to wake Ralph in order to have his company, and the other is the desire to retrieve her glasses. According to conditions (1-C) and (2-C) as Sarah enters the room she both intends her entering the room to enable her to obtain the glasses and remembers her previous desire to obtain the glasses and understands her present action as satisfying her desire. Clarke raises the question whether we should automatically assume that her intention is authoritative in explaining her action? Suppose in fact, Sarah's conscious intention does not play a causal role in bringing about her entering the room but that her desire to wake Ralph does. In such a case, we would deny that retrieving her glasses was the real reason she entered the room and would judge that she was merely fooling herself (Clarke 2003, p. 22). This seems to show that conditions (1-C) and (2-C) are not sufficient for explaining action.

In reply to this, Ginet cites his opposed intuition:

In response to the first claim I can say only that I have the strong contrary intuition that, even if Sarah's antecedent desire to get her glasses did not cause her action of entering, the fact that that act was accompanied by her intention that by that act she would satisfy that desire (or contribute to satisfying it) does entail that her aim in entering was to satisfy that desire, that she entered in order to satisfy it, that the desire was a reason (though not necessarily the only reason) for which she entered. And it would simply be question-begging to argue against my intuition by saying that the desire cannot be a reason for which she performed that action if it played no role in causing the action. (Ginet 2008, p. 231)

On the one hand, Ginet is willing to concede that there may be more than one reason for which Sarah entered the room; on the other hand he dismisses the appeal to a causalist explanation as question-begging.

One reason for rejecting Ginet's reply rests on the previously discussed principle of *non-transparency with respect to the causal-efficacy of reasons* (NTCER). According to this empirically well-grounded principle, agents are sometimes introspectively ignorant as to the relative causal strength of their reasons. In an extreme case, an agent might suppose that she is acting on the basis of a certain reason when it is causally inert with respect to bringing about the relevant action. This is likely to be true in cases where the consciously ascribed reason is an *excuse* which masks the actual, causally efficacious reason. Given this possibility, it may very well be the case that Sarah in no respect acts in order to retrieve her glasses but acts solely in order to wake Ralph. Yet even if we concede that Sarah partly acts in order to retrieve the glasses as Ginet suggests, it will still be the case that the primary explanation of why she entered the room will be in order to wake Ralph. Conditions (1-C) and (2-C) will at best be a supplement to the explanation of the action.

The second problem with Ginet's defense pertains to his charge of question-begging. The reason the causalist is innocent of this charge is because the noncausalist account of action is insufficiently explanatory. Whenever an event needs to be explained, what is sought is an identification of the cause or causes of the event. Consider the event of a window breaking. Assuming that this is an agent-involving event, one might choose to explain it noncausally by saying that it was broken because A didn't have her keys. But while this is a beginning to an explanation of how the window got broken, it is incomplete. To provide a full explanation one has to specify the immediate *cause* of the breaking, say, a rock thrown by the agent impacting the glass. This paradigm applies to events in general. But perhaps the noncausalist will insist that explanations in terms of events do not apply in cases of agents' actions. She may say that actions need to be explained in terms of intentions only, not desires or other causes. This move is unsatisfactory, however, because it conflicts with basic considerations concerning mental causation. As discussed previously, simply by virtue of their supervenience on physical states, mental states participate in the causal commerce of the world. By way of their physical bases, mental events cause other mental events as well as physical events such as actions. So in asking for an explanation of an action, one automatically seeks a causal explanation of it. It is inadequate to cite intentions understood noncausally because intentions, qua mental events inherit causal roles from the physical states they supervene on.

A further objection to Ginet's defense (Clarke 2003, pp. 22-23) may be summarized as follows. When Sarah decides to enter the room she decides for a reason, R_1 . Suppose that reason either is a desire or corresponds to one. Then,

according to Ginet, it appears that she acts on that desire only if she has a second-order intention, R_{-1} to satisfy R_1 . This is in fact what condition (1-C) seems to require. Then, in order to act on R_{-1} Sarah will have to have the desire R_{-2} to do so. As a result, in the absence of a causal explanation, which, of course, is unavailable to the noncausalist, an infinite regress is generated. Clarke then presses this concern further by saying:

But one can act for a certain reason, and citing a desire can provide a reason-explanation of one's action, even if one does not have when one acts any such second-order intention. Sarah, for example, might act on her desire to get her glasses even if her only intention when she enters the room is to retrieve her glasses. And if such a concurrent intention is not necessary, then the question arises what, in its absence, suffices as a response to the challenge described earlier. (Clarke 2003, p. 23)

Clarke's point is well taken. Second-order intentions do not figure in all actions. Sometimes we act on the basis of immediate desires. But condition (1-c) seems to require that all actions be accompanied by second-order intentions.

In response, Ginet denies that an infinite regress is entailed.

Let us assume for the moment that we may say that Sarah had a reason for deciding to enter the room and that this reason was her desire to get her glasses. My account does not in fact imply that this can be true only if concurrently with her deciding Sarah intended of that act of deciding that it contribute to her getting her glasses. My account claims only that an accompanying intention with the content "This act is to satisfy that desire" is a sufficient condition for the truth of "The agent performed the act in order to satisfy that desire", not that it is necessary. (Ginet 2008, p. 232)

In light of this explanation, condition (1-C) was misinterpreted as a necessary condition when in fact it is a sufficient one. When Sarah decides to enter the room she does not have to have an intention R_{-1} to satisfy the original intention R_1 . All that Ginet is claiming is that *if* R_{-1} is present, then it is true that the agent performed the act in order to satisfy

that desire. I must concede that this reply effectively dispenses with the threat of an infinite regress, though the other concerns about the noncausalist account remain.

Conclusion

The principle of non-transparency (NTCER) refutes the notion that the presence of the second-order intention is sufficient to explain the action. As discussed above, Sarah make consciously intend to retrieve her glasses as she enters the room, yet it may still be the case that the real reason on which she acts is to wake Ralph. Conscious intentions, whether first or second order, are defeasible explanations of action. An additional concern for the noncausal explanation of action is that noncausal explanations are incomplete. They only tell us part of the reason why an action takes place. Finally, the assumption of the supervenience of the mental on the physical implies that mental events, such as intentions inherit causality and causal efficacy from their physical bases thereby precluding a noncausalist explanation of action.

CHAPTER 4 EVENT-CAUSAL LIBERTARIANISM

Introduction

Event-causal theories comprise another family of libertarian theories. These theories argue that under the right conditions, some agent-involving event nondeterministically causes the free actions of a human agent. Typically, such theories take advantage of the general conditions for a free action that are specified by compatibilist theories, and then add an indeterministic causal factor. Important event-causal theorists include Daniel Dennett, Laura Eckstrom, Alfred Mele, Robert Nozick, Mark Balaguer, Peter van Inwagen, and Robert Kane. In the present chapter I will explicate the salient details of most of these theories and explain the ways in which they individually fall short. Among these theorists, Robert Kane is the most influential and widely discussed. For this reason I will dedicate the entirety of the subsequent chapter to criticizing his theory.

Early Event-Causal Theories

In *Brainstorms* Daniel Dennett articulates an early event-causal theory without endorsing it. The central core of the event-causal view that he puts forward for examination is as follows:

The model of decision making I am proposing has the following feature: when we are faced with an important decision, a consideration-generator whose output is to some degree undetermined produces a series of considerations, some of which may of course be immediately rejected as irrelevant by the agent (consciously or unconsciously). Those considerations that are selected by the agent as having a more than negligible bearing on the decision then figure in a reasoning process, and if the agent is in the main reasonable, those considerations ultimately serve as predictors and explicators of the agent's final decision. (Dennett 1978, p. 295)

In support of this model, Dennett argues that it presents intelligence as the difference maker in free action, that it is expedient since we are often pressed for time when deciding how to act, and that it explains how we can be the author of our decisions. We are the author of our decisions on this model because it is we who decide which considerations should bear on our ultimate decision. Dennett also contends that the model locates indeterminism at just the right place.

It would be insane to hope that after all rational deliberation had terminated with an assessment of the best available course of action, indeterminism would then intervene to flip the coin before action. It is a familiar theme in discussions of free will that the important claim that one could have done otherwise under the circumstances is not plausibly construed as the claim that one could have done otherwise given exactly the set of convictions and desires that prevailed at the end of rational deliberation. So if there is to be a crucial undetermined nexus, it had better be prior to the final assessment of the considerations on the stage, which is right where we have located it. (ibid.)

The theory thus postulates that the indeterminism occurs somewhat early in the production of action, before deliberation has terminated.

Critics of this theory have stated that limiting indeterminism to the coming to mind of considerations is too weak to ground freedom of action. Laura Ekstrom highlights the fact that apart from the indeterminism of the coming to mind of considerations, the rest of the action production process is determined. She says that a view such as this leaves unanswered “why an act that is the causally necessary outcome of whatever considerations have happened to occur is plausibly claimed to be the originated by the agent” (Ekstrom 2000, p. 137). This directly attacks Dennett’s claim that this model shows us to be the authors of our actions. Suppose I am deciding whether to move to a different apartment. Then on Dennett’s account, a variety of considerations will indeterministically come to mind, some relevant to my decision, some not relevant. But

this is the only aspect of the decision-making process that is undetermined. The particular weights I place on each consideration will be determined, along with the issuance of the action once the selection of considerations is made. This situation is comparable to having an automated process produce an action, once an automated selection is made from among a randomly generated set of considerations. The selection from among the considerations is unfree because the weights are predetermined factors of the agent's constitution. Moreover, since there is no further indeterminism in the production of action, once the selections are made, the decision is computed deterministically. These facts undermine the notion that the action is originated by the agent. It is also significant that the set of considerations which are indeterministically produced include both irrelevant as well as irrelevant ones. This suggests that they are random productions, and thus not *authored* by the agent.

The indeterminism of the coming to mind of considerations is not only too weak to ground free action, but it appears to be located too early. Dennett thought that it would be insane to place the indeterminism after deliberation had concluded. But it could be argued that agents would prefer to be able to continue to make revisions even after reaching a provisional decision. Even if I decide to move to the Windmeadows Apartments, I should be able to reevaluate and possibly alter my choice all the way up until I sign the lease. A related consideration pertains to *akrasia*. If weakness of will is to be possible, as many philosophers think it is, it must be possible for me to fail to act even after making a definitive decision. This implies that some indeterminacy needs to exist all the way up until the moment of action.

Another criticism of Dennett's theory is put forward by Randolph Clarke. He argues that on Dennett's account, agents are not difference-makers in the right way. He argues that on the view presented by Dennett

agents might be said to make a difference between what happens but might not have and what does not happen but might have, but such a difference is made in the occurrence of something nonactive or unfree prior to the action that is said to be free, not in the performance of the allegedly free action itself. Failure to secure for directly free actions this libertarian variety of difference-making constitutes a fundamental inadequacy of deliberative libertarian accounts of free action. (Clarke 2003, p. 64)

On Dennett's model, the difference-making is with respect to a selection from among considerations that come indeterministically to mind. This stands in contrast to a more robust conception of an agent making a difference in the very performance of a free action. On the latter conception, some things happen that would not have happened, and some things do not happen that would have happened, just because of something the agent voluntarily does. But on Dennett's model, actions are produced by a causal chain in which each stage, save the particular considerations that come to mind, is the deterministic outcome of its immediate antecedents. Thus Dennett's theory falls short again for similar reasons as those adduced by Ekstrom. Let us now turn our attention to another early event-causal theory, that of Robert Nozick.

As with Dennett, Nozick does not fully endorse this proposal; he merely offers it for examination. He starts by characterizing the phenomenology of deliberation as one in which we mull over reasons and decide which among them have greater weight. He further argues that we do not *discover* the weights in this manner. Rather, it is through deliberation that relative weights are assigned.

The reasons do not come with previously given precisely specified weights; the decision process is not one of discovering such precise

weights but of assigning them. The process not only weighs reasons, if (also) weights them. At least, so it sometimes feels. This process of weighting may focus narrowly, or involve considering or deciding what sort of person one wishes to be, what sort of life one wishes to lead. (Nozick 1981, p. 294)

This is a plausible description of what goes on in deliberation. Surely when I consider whether or not to move to a new apartment I review various reasons and assign relative weights to them. Then, if all goes well, I base my decision on the weightiest reason or set of reasons. But this immediately raises the possibility that the respective weights reflect predetermined features of my constitution. As such, they would not contribute to my decision of act being free. For example, if antecedent conditions deterministically cause me to assign the greatest net weight to the cost of an apartment, this weighting is part of the background of determinism rather than a freedom conferring feature.

Fortunately, Nozick has solution to this problem.

First, the decision may be self-subsuming; the weights it bestows may fix general principles that mandate not only the relevant act but also the bestowing of those (or similar) weights. The bestowal of weights yields both the action and (as a subsumption not a repetition) that very act of bestowal. For example, consider the policy of choosing so as to track bestness: if the act weren't the best you wouldn't do it, while if it were best you would. The decision to follow this policy may itself be an instance of it, subsumed under it. (Nozick 1981, p. 300)

Nozick's ingenious move is to make the freedom of the decision derive from the decision itself. This is accomplished by the putative self-subsuming character of the assignment of weights. Suppose I am deciding to move and consider various reasons including their respective costs, aesthetic qualities, distance from work, etc. As it turns out, I assign the greatest net weight to the cost factor. On Nozick's account, this assignment of weight is not a predetermined feature of my constitution; rather, it originates with the decision itself as a self-subsumption. There is no need to look to the

past for the explanation of the weight assignment; the assignment occurs as an intrinsic feature of the choice.

Assuming for the moment that the self-subsumption maneuver is intelligible, one may still wonder whether it renders the assignment of weights as random. Since nothing outside the decision itself determines the weights, what are the weights based on? If the answer is that they are not based on anything, then the bestowal of weights appears random and irrational. Nozick's response to this is to draw an analogy to explanatory laws in general. he says of the assignment of weights that "It will no more be a random brute fact that is the holding of a fundamental, deep explanatory law that subsumes and thereby explains itself" (Nozick 1981, p. 301). This model of explanation has the following form:

P: any lawlike statement having characteristic C is true.
P is a lawlike statement with characteristic C.
Therefore P is true. (Nozick 1981, p. 119)

An example intended to show self-subsumption in action is:

S: Every sentence of exactly eight words is true.
S has exactly eight words.
Therefore S is true. (ibid.)

If the analogy between self-subsuming explanatory laws and the bestowal of weight when making a decision holds, then we have a good reason for seeing the bestowal of weights as non-random and rational.

Timothy O'Connor, however, argues that the analogy does not succeed (O'Connor 2000, p. 31-32). Leaving aside the soundness of Nozick's account of self-subsuming explanatory laws, he says that explanatory laws are unlike the bestowing of weights in that the former do not *become* true in time, whereas the latter is an act that is realized only when the bestowal of weights occurs. Consider Nozick's principle that the

very sentence with exactly eight words is true. If that were a true proposition, then it would be timelessly true. By contrast, on Nozick's account, one's reasons only acquire the weight they do at a particular time, namely, when the weights are assigned at the moment of decision. But if the analogy between self-subsuming explanatory laws and weight assignments breaks down, there "doesn't seem to be a basis for claiming that a decision can be explained (solely) by reference to its own content" (ibid.)

O'Connor's objection may be better appreciated by examining the analogy in detail. Although Nozick does not explain how the analogy between a self-subsuming explanatory law and the assignment of weights is to work, it seems to be as follows. When I bestow the greatest net weight on the cost of the apartments I operate on the principle "Lower cost is the most important consideration in deciding between apartments." Next comes the empirical claim that "Windmeadows Apartments are judged to be of lower cost." And from this we get the instantiation of the general principle, "The lower cost of the Windmeadows Apartments is the most important consideration in deciding between apartments." Nozick's point seems to be that the judgment that the lower cost of the Windmeadows Apartments is of the greatest importance is seen to be rational, and thus not random, because it is an instance of the general principle that lower cost is the most important consideration. Moreover, he may be arguing that the logical connection between the principle and its instantiation holds even though they are co-occurrent. But the only thing that makes the general principle true is the fact of assigning the greatest net weight to cost at a particular point in time. There is no self-subsuming at all. Rather, the general principle merely *explains* why the lowest cost is the most important consideration in the present instance. Furthermore,

the general principle does not in any sense explain itself; one must look to the agent's volition for its validity. Thus in the absence of an explanation as to why the weights are assigned as they are, the assignments appear random and non-rational.

A related objection to Nozick's strategy is presented by Clarke:

On the suggested view, the assignment of weights to reasons is part of the content of a free decision; that particular weighting is (among) what is chosen. The assignment of weights, then, is intrinsic to the decision. Thus, what is being claimed is that an explanation of the event can be provided simply by citing some intrinsic feature of that very event. And this is just incredible. (Clarke 2003, p. 51)

As he goes on to say, in order to explain the beating of hearts and the formation of stars, etc., we must look to causes of the respective events, not their intrinsic features. On Nozick's proposal, the freedom of the decision derives from the very bestowal of weights that is constitutive of the decision. But such an account is circular. Nozick attempts to break out of the circle by invoking self-subsumption. But, as we have seen, we have to look outside mere fact of the bestowing of the weight for an explanation of the bestowal. Rather than looking to the intrinsic fact that the decision is an act of bestowing, we must turn to the volition of the agent to explain the bestowal. Thus Nozick's proposed event-causal theory cannot be accepted.

Van Inwagen's Account

The next theory to be examined is put forward by Peter van Inwagen in his influential 1983 book, *An Essay on Free Will*. There he presents a relatively straightforward desire-belief (DB) model of human action on which agents act in pursuit of their desires to the extent that they believe the action in question will satisfy those desires. In addition, on this particular libertarian theory although an agent's desires and beliefs cause his actions, they do not determine, or necessitate, those actions. Van

Inwagen discusses an example in which a thief has to consider whether to give in to the temptation to rob a poor-box, or to refrain from stealing on the basis of an alternative desire, namely the promise he had made to his dying mother. As it happens, the thief manages to restrain himself.¹ Van Inwagen says:

If the model of the causes of action we are considering is correct, his refraining from robbing the poor-box (R) was caused but not necessitated by his desire to keep the promise he made to his dying mother coupled with his belief that the best way to do this would be to refrain from robbing the poor-box (DB). Let us suppose that the second [i.e., the present] model is correct: R was caused by DB and DB did not *have* to cause R; it just *did*. We may suppose that God has thousands of times caused the world to revert to precisely its state at the moment just before the thief decided not to steal, and has each time allowed things to proceed without interference for a few minutes, and that DB caused R on about half these occasions. On the other occasions, we may suppose, DB did *not* cause R: instead the thief's desire for money, coupled with his belief that the best way for him to get money was to rob the poor-box. (van Inwagen 1983, pp. 140-141)

The possibility of God resetting the circumstances and replaying it numerous times is significant. Suppose the event is replayed a large number of times and the outcome is the same on each occasion. If this were the case it would be virtually undeniable that the action (or the refraining from action) flowed from the will of the would-be thief, from his desires and belief. There would be no doubt that he was the source of the refraining, that the refraining was up to him. However, van Inwagen supposes that on around half of the replays the thief did not hold back from robbing the poor-box. Given this consideration, one may wonder whether the agent (sometimes a thief, sometimes not) is really in control of what he does. Although DB produces the refraining on half of the

¹ The fact that the action in question is really a refraining from action rather than a positive action should be irrelevant to the issues under discussion.

runs, it doesn't do so on half of the runs.² There is thus doubt about the degree of control possessed by the agent. Pressing this concern, O'Connor writes:

Even though the causal indeterminist [event-causal libertarian] account allows for the real possibility of different courses of action, any of which would be "controlled" by the agent in the minimal sense of being an "outflowing" of the agent, it's not "up to the agent," something he "has a choice about," just which potential cause will be efficacious in any given instance and so which action will actually occur. It is, rather, a matter of its falling under a statistical or quasi-statistical tendency that governs the general pattern of behavior of types of circumstances over time, and this probabilistic tendency clearly is not something the agent has any choice about. (O'Connor 2000, p. 29)

The same agent, with the same background and the same desires and beliefs, sometimes does one thing and sometimes another. Since his desires and beliefs are the same in both cases, what he does doesn't appear to be up to him. Rather, which action he performs seems to be a matter of a "probabilistic tendency". Neither the agent robbing the poor-box or refraining from doing so seems to be up to the agent; consequently, he does not appear to be free.

Van Inwagen is aware of this objection. He puts it forward as one that he, as a libertarian, is committed to rejecting.

"If an agent's act was caused but not determined by his prior inner state, and if nothing besides that inner state was causally relevant to the agent's act, then that agent had no choice about whether that inner state was followed by that act" (van Inwagen 1983, p. 149).

This proposition essentially encapsulates the above objection by O'Connor. The inner state is the particular configuration of desires and beliefs. On this particular libertarian theory, the desires and beliefs do not *determine* the actions. But on the assumption that a different action results on about half the runs, it would appear that the agent "had no

² This is a version of the "luck objection" we encountered earlier.

choice about whether that inner state was followed by that act.” This is the natural conclusion to draw. It is so natural that van Inwagen himself admits that it is “puzzling that this proposition should be false” (ibid.). But despite this admission he goes on to say that it must be rejected unless one is to conclude, among other things, that free will is not possible. Since he finds the latter *inconceivable*, he opts for rejecting the above proposition (ibid.). This is not a satisfactory reply in discussing a controversial topic to which there are partisans on either side.³ A more satisfactory defense to the objection is presented by Clarke. Since the agent sometimes robs the poor-box and sometimes doesn’t, it may be said that we have to wait and see what the agent does. To this Clarke replies:

But, of course, although we can only wait and see, this is not all that *the thief* can do; he can (at least, it is open to him to) make up his mind. And when he does—when *R* occurs—it is not *merely* a matter of chance whether *R* occurs, if that means no one exercises any active control over the occurrence of *R*. (Clarke 2003, p. 100).

On this response, the agent has a kind of initiatory power to decide what to do, given his beliefs and desires. Thus this may be seen as an amendment to van Inwagen’s position as “making up his mind” plays no role in the original account. At first glance, this initiatory power reduces the degree to which what the thief does is a matter of chance. But on further inspection, the reply only pushes the problem back a step. We can still ask what happens on a large number of replays of the scenario. Instead of looking to see whether the agent’s desires and beliefs cause him to refrain or not, we can ascertain whether he makes up his mind to rob the poor-box or not. Making-up-one’s-mind is a mental act. If the thief refrained on all the runs (or perhaps on a great majority

³ Although the reply is not satisfactory, presumably it is stated in earnest. This shows the degree to which the presumption of free will is entrenched in human thought.

of them) we would be sure that he was in control of his decision, i.e., his mental act of making up his mind. But if he makes up his mind each way on only half of the runs, the same doubt arises as did in relation to his desires and beliefs. It will appear that it is not up to him what he decides, that what he does is just a matter of a statistical tendency. Thus van Inwagen's version of event-causal libertarianism cannot be accepted.

Recent Event-Causal Theories

In 2000 Laura Ekstrom put forward a new event-causal theory based on the formation of preferences. For Ekstrom, a *preference* is “a specifically processed desire—a desire formed by a process of critical evaluation with respect to one’s conception of the good” (Ekstrom 2000, p. 113). A key consideration is that these preferences are generated indeterministically. It is also significant that this is the only locus of indeterminism in her account. The notion of preferences is central to her conception of the *agent* as an evaluating and choosing faculty together with an aggregate of preferences and acceptances (ibid.). Next, a preliminary to action is the agent’s *intention* which is “agent caused just in case it results from a normal causal process from a preference for acting as specified in the content of the intention, where the preference itself is the output of an uncoerced exercise of the agent’s evaluative faculty, the inputs into which cause but do not determine the decisive formation of the preference” (Ekstrom 2000, p. 114). This element of the theory supplies a defense against freedom undermining factors similar to those found in compatibilist theories. Finally we arrive at the definition of *free action* itself “An action is free only if it results, by a normal causal process, from a pertinent intention (e.g., an intention to perform the act here and now) that is *caused by the agent*, where this latter term (*‘caused by the agent’*) is reducible to event-causal terms” (Ekstrom 2000, p. 115). The qualification, “reducible

to event-causal terms” distinguishes this theory from agent-causal accounts which make the agent alone the sufficient cause of action. On the present theory, the agent’s causing of the action is mediated by an array of events, including mental states for example.

In accordance with Ekstrom’s model, a decision to move to a new apartment complex is a free action if it expresses the intention to act on an indeterministically formed, uncoerced preference. While it is certainly an asset to the theory that the preferences are uncoerced and indeterministically formed, one may wonder if this is sufficient for freedom. Clarke notes that on Ekstrom’s view:

we have undetermined actions — the formations of preferences — among the causes of free decisions. But she does not require that these preference-formations either be or result from free actions. Nor can she require this. Any free action, she holds, must be preceded by a preference-formation. An infinite regress would be generated if these preference-formations had to either be or result from free actions. (Clarke 2003, p. 63)

Is this criticism fair? Ekstrom explicitly defines free action as action that expresses undetermined, uncoerced preferences. Is it fair to demand another stratum of freedom on which to ground free actions as she has defined them? Ekstrom can argue that the indeterministically formed preferences are constitutive of the identity of the agent. The actions based on these preferences will then truly express who the agent is. Moreover, the agent will appear to be the originator of the action. But what about the fact that every facet of the agent except for preference-formation is determined? This general observation seems to undermine the opinion that the agent is free. Furthermore, the agent is not responsible for her preferences in an active sense. She can’t be by virtue of infinite regress as Clarke notes. The preference-formations are agent-involving events, but not actions of the agent. Consider the agent who indeterministically forms the

preference to move to where the cost is lower. This preference-formation seems closer to a spontaneous emergence from within the person, than an action by her. Moreover, every other facet of the person prior to the preference-formation, as well as everything that comes after the formation of the preference—is determined. On the whole, we get a picture of an unfree person with spontaneous eruptions of particular preferences.

The moral of the story seems to be that indeterminism in itself does not generate freedom. Indeterminism does have the effect of screening off the deterministic influences from the past. Since, from a libertarian perspective, deterministic influence is freedom undermining, screening off such influences may seem to promote freedom. But freedom is distinct from indeterminism; it is more like an active power. Merely finding a way to introduce indeterminism into the production of action does not introduce freedom into it. Indeterminism is only the absence of determinism whereas freedom is a kind of self-determination or self-production that is independent of other influences. Ekstrom's theory, as well as the previous event-causal theories we've examined, introduce indeterminism into the production of action, but don't show that the agents are thereby free.

Mark Balaguer recently articulated another event-causal libertarian theory. Its central feature is what he terms a "torn decision".

A torn decision is a decision in which the person in question (a) has reasons for two or more options and feels torn as to which set of reasons is strongest, i.e., has no conscious belief as to which option is best, given her reasons; and (b) decides without resolving this conflict—i.e., the person has the experience of "just choosing". (Balaguer 2004, p. 382)

These torn decisions are held to be free because on the one hand they are undetermined, and the other they are appropriately non-random. They are undetermined because they are apparently not determined by the state of the world up

until the decision. That is to say, the state of the world seemingly does not select which of the options will ultimately be decided on. In addition, torn decisions are non-random in that the choice that is made is acted on for reasons. This follows from the fact that there is at least one reason for each of the options.

While the latter consideration seems true, one may wonder about the former. It might be argued that the ultimate decision is not truly undetermined because it is the outcome of a deterministic process in the brain. That is, the decision may be torn from the standpoint of introspection, but at a deeper level which option gets selected is because of deterministic causation. Balaguer anticipates this objection and offers the following argument.

On the view I have in mind, our torn decisions aren't determined by anything: they're not determined by our reasons, and they're not determined by the physical states of our brains just prior to the moment of choice; rather, we just make them. If the token-token identity theory of decisions is right, and if our torn decisions are undetermined at the moment of choice in the way described above, then we ought not to say that these decisions "arise out of" neural events; what we ought to say is that they are neural events (and that they are undetermined). But while they are neural events, if the token-token identity theory is true, they are also mental events; in particular, they are purposeful, intentional decisions; Ralph, for instance, decided to move to New York consciously and on purpose, fully aware of what he was doing and in full possession of his faculties. Thus, it doesn't follow from the mere fact that our torn decisions are physical events that they're unfree, or random in some appropriate sense; for given the token-token identity theory, our torn decisions are still mental events, and in particular, they are still conscious, intentional, purposeful, and so on. (Balaguer 2004, p. 387)

There are two main claims to consider in the above. One is whether the token identity theory renders the neural-mental pairs as indeterministic. Balaguer seems to be arguing that merely because each of the relevant neural states is also a mental state, that by virtue of being mental they are undetermined. But this is not the most natural conclusion to draw. What the token identity theory directly shows is that whatever properties a

particular neural event has is also shared by its correlative mental event. That is, if the neural events are shown to be determined, then by virtue of the token identity theory the mental states which map onto them would also be determined. Suppose a certain pattern of neural states persists for fourteen seconds. Suppose also that the pattern of neural states is token identical to an experience of pain. What we must then conclude is that the pain experience persisted for fourteen seconds. The same consideration applies to the pattern of emergence of neural events and mental events. If certain neural events proceed deterministically, then so must the mental events that are token identical to them. Matters could be different, however, if a suitable brand of dualism were true. If mental events were token distinct from neural events, then they would not have to share the same properties as any particular set of neural events. On that assumption, Balaguer's claim that the mental events are undetermined could hold true, but his theory would be radically different in character.

The second claim Balaguer appears to be making is that in spite of token identity with physical states, mental events are nonetheless free because they are conscious, intentional, and purposeful. He is certainly correct in saying that mental events can be conscious, intentional, and purposeful even if they are token identical to physical states. But this need only mean that the mental properties in question supervene on the underlying neural base without in any sense cancelling or weakening the physical characteristics of the neural events. In other words, the fact that a certain pattern of neural states instantiates the properties of being conscious, intentional, and purposeful, does not in any way imply that it does not also instantiate the property of being causally determined. Rather, what token identity theory shows is that agents who are conscious,

intentional, and purposeful can also be causally determined. Against this, it may be argued that certain patterns of intentionality and consciousness are incompatible with causal determinism. For example, it might be said that one could not act consistently with one's purposes unless those purposes were the basis of one's actions. But what token identity shows is that either no pattern of mentation that is inconsistent with the physical base will actually come about, or that one's phenomenology will always represent one's mental states in a way that does not conflict with what one knows about one's neural state(s). Contrary to Balaguer, what token identity does is compel us to reevaluate the status of our mental states in the light of the corresponding neural states, not to put aside the physical properties of the mental-neural pairs, such as their proceeding by a deterministic causal process.

Daring Soft Libertarianism

The final event-causal theory we will examine at the present time is one put forward by Alfred Mele in 2006. Although he discusses several theories, the most innovative is *daring soft libertarianism* (DSL).

My daring soft libertarians especially value a power to make decisions that are not deterministically caused—a certain *initiatory* power. They are causalists about actions, including decisions and...they believe that agent causation leaves the problem of luck intact. Not wanting to be saddled with a questionable species of causation for which they have no use, they opt for event-causal soft libertarianism... [T]hey believe that although people sometimes try to bring it about that they decide to *A*, people never try to decide to do what they judge best to do while also trying to decide instead to do what they are tempted to do. According to their causal theory of action, actions are events with the right sort of causal history, decisions are made for reasons, and deciding for a reason *R* to *A* requires that *R* (or the agent's apprehension of *R*) play a part in causing the decision. (Mele 2006, p. 113)

In addition to the elements indicated above, the general features of DSL may be summarized as follows.⁴ First, it requires only non-robust alternatives when faced by Frankfurt scenarios. Frankfurt scenarios are those in which an agent will either perform an action on her own or be compelled to do so. Having a merely non-robust alternative means that she cannot do otherwise than A at *t*. The only alternative she has is to A freely or by force.⁵ DSL also situates indeterminism in the production of action between the decision to intentionally A and the corresponding action. It thus situates the determinism at a later stage than the formation of preferences or the coming to mind of considerations. According to DSL, actions that are themselves deterministically caused can inherit freedom from previous actions that are suitably related. Finally, the motivation for DSL is that its adherents perceive that indeterminism affords a more desirable species of free action and moral responsibility than determinism. This last feature will become relevant when the assessing the theory as a whole.

As a major selling point of DSL is that is said to be able to “stare down” what is known as the luck objection to libertarianism we should turn now to Mele’s formulation of the argument. Assume the standard definition of *determinism* as the thesis that at any instant exactly one future is compatible with the state of the universe at that instant and the laws of nature. Let libertarianism (L) be understood as comprising three claims (a) that some human beings sometimes perform basically free actions, (b) that some agents are basically morally responsible for their actions, and (c) that incompatibilism is true. A *free action* is one such that when the agent performs it, it either was at that time

⁴ This summary derives from the analysis supplied by Maurer and Gutermann (2007).

⁵ By contrast, a robust alternative would be if she could have done A at *t* or otherwise than A at *t*.

possible for the agent to have done otherwise, or else if the agent was not able to have done otherwise at that time, she nevertheless freely performed some earlier relevant act of which she could have done otherwise. For example, consider someone driving off the road at t . If the person could have done otherwise at t than drive off the road, then the act of driving off the road is a free act. But suppose person's motor coordination is impaired through intoxication at t so that the person cannot voluntarily stay on the road. On this definition of free action the act of driving off the road at t will still count as a free action provided that the agent could have done otherwise than become intoxicated earlier. As for *basically free* and *basically morally responsible*, Mele spells these out as follows:

Basically free actions are free A-ings occurring at times at which the past and the laws of nature are consistent with the agent's not A-ing then. A-ings for which an agent is *basically morally responsible* are A-ings for which he is morally responsible that occur at times at which the past and the laws of nature are consistent with his not A-ing then (Mele 2006, p.114).

We are now ready for the reconstruction of Mele's version of the luck argument.⁶

1. According to L, an agent's A-ing at t is basically morally responsible if the past and the laws of nature are consistent with the agent not A-ing at t .
2. Suppose an agent in W_1 basically freely As at t .
3. According to L, agents are basically morally responsible for their basically free actions.
4. An agent's A-ing at t is basically free if the past and the laws of nature are consistent with the agent not A-ing at t .
5. Suppose an agent in W_1 As at t .
6. Then it follows that in W_2 the agent does not A at t .
7. But W_1 and W_2 are identical up until t .
8. Likewise, the agent's properties are identical in W_1 and W_2 .
9. Thus it is a matter of luck that the agent As in W_1 but does not A in W_2 .
10. Since it is a matter of luck that the agent As in W_1 but does not in W_2 , then the agent is not basically morally responsible for A-ing.
11. Therefore L is false.

⁶ This reconstruction is based on Mele, 2006 pages 6 through 9.

Mele believes that DSL can “stare down” this objection:

Daring soft libertarians (DSLs) try to stare down the problem of present luck. They claim that present luck [i.e. the compatibility of a significant act’s nonoccurrence with the immediate past and laws] is entailed by an agent’s having a kind of initiatory power that they value and that its presence in a case of action does not preclude the action’s being freely performed or the agent’s being morally responsible for it. The softness of their libertarianism makes their situation less treacherous than that of [traditional] libertarians. Soft libertarians do not assert that free action and moral responsibility require the falsity of determinism. (Mele 2006, p. 113)

As a version of “soft libertarianism”, DSL doesn’t insist that free action and moral responsibility are incompatible with determinism. It merely maintains that indeterminism allows for a more desirable species of free action and moral responsibility. This raises a question as to the general viability of DSL as a libertarian theory. Most libertarians are committed to the view that free will and moral responsibility *are* incompatible with determinism. The liberality of DSL places it close to compatibilism. Perhaps the one element that warrants counting DSL among libertarian theories is that it does claim that a species of free agency that is incompatible with determinism is realized in our world.

Mele’s central defense against the luck objection rests on the capabilities of “little agents”.

Consider the first time a normal child, Tony, makes a decision about whether to snatch a toy from his younger sister. [. . .] Tony knows his father is nearby; and, on the basis of some unpleasant experiences, he associates taking the toy with his sister’s screaming and his father’s scolding him. He decides not to snatch it and feels a little frustrated. [. . .] Suppose now that owing to Tony’s being an indeterministic decision maker and to his being tempted to take the toy, there was a significant chance at the time that he would decide to take it. In another world with the same past and laws of nature, that is what he decides to do, and he proceeds to grab the toy (with predictable results). Does that entail that Tony has no moral responsibility at all for deciding not to take the toy? (Mele 2006, pp. 129–130)

The expected answer to the latter rhetorical question is “No.” But this reply is dependent on a background that assumes responsibility for actions in general. However, if we take the luck objection seriously, we are calling these background assumptions into question. Quite naturally, if generally speaking agents are morally responsible for their actions, then a child would bear some modicum of responsibility for her actions, despite her immaturity. But the luck objection contemplates a mature agent, and shows that such an agent, who is normally thought of as morally responsible and free, in fact does not exert sufficient control over her decisions to be considered genuinely free. Consequently, when confronting the example of a little agent, the same luck objection applies. In an alternative world, Tony would fail to restrain himself and would snatch the toy from his sister.

Mele continues his resistance to the luck objection.

Moral responsibility is very commonly and very plausibly regarded as a matter of degree. If young children and adults are morally responsible for some of what they do, it is plausible. . . that young children are not nearly as morally responsible for any of their deeds as some adults are for some of their adult deeds. When we combine our recognition of that point with the observation that the good and bad deeds of young children are relatively trivial in themselves, we should be struck by the implausibility of stringent standards for deserved moral praise and blame of young children – including standards the satisfaction of which requires the absence of present luck. [. . .] [A]s the frequency of the indeterministically caused free actions of little agents increases and as the range of kinds of situations evoking such free actions expands, the agents take on greater moral responsibility for associated [probabilities of future action] and for their morally significant free actions. This. . . helps to account for the fact that the moral credit and blame that little agents deserve for their indeterministically caused free actions tend to increase over time. (Mele 2006, pp. 131–132)

This is a potent observation. The denier of free will is obliged to say why she distinguishes among agents with respect to degrees of responsibility. The answer is that some aspects of our traditional notions of moral responsibility can be justifiably retained

even on the assumption that free will does not exist. More precisely, certain notions having a family resemblance to “true” or “real” moral responsibility would survive in a deterministic world. For example, it seems clear that some forms of non-moral responsibility would be possible on either compatibilist or incompatibilist assumptions. For example, it can be true to say that someone is responsible (in the sense of being the cause of) taking an item from a store without paying. It also seems plausible that it would be in society’s interest to prevent such behavior. Accordingly, upon identifying someone as having stolen something, it may be desirable to impose corrective sanctions (e.g., rehabilitative conditioning) in order to alter the proclivities of the person. Moreover, this could be justified on incompatibilist grounds, as it does not require assigning unqualified blame to the individual. For even if it could be shown that the agent could not have done otherwise than steal, it would still be helpful to identify her as the thief, and to bring it about that she does not steal in the future. With respect to young children, exhibiting a connection between behavior and its results in certain cases will help to train the child to behave in preferred ways. My contention is that many such quasi-moral notions and practices could be preserved if hard determinism were true. In general, institutions and practices that employ an attributivist notion of responsibility and which serve an instrumental function could be retained. These include sanctions to influence behavior and communications of approval or disapproval. Likewise, the cultivation of the reactive attitudes could be justified insofar as they serve an instrumental function. For example, since the reactive attitudes are the motivational basis for our taking and assigning blame and responsibility, those moral practices that depend on them can and should be retained by the hard determinist.

The above considerations plus the applicability of the luck objection itself to little agents seems to adequately rebut Mele's objection to the luck objection. As a result, the luck objection itself undermines daring soft libertarianism. We are not quite finished with event-causal libertarianism, however. We still need to examine the influential theory put forward by Robert Kane.

CHAPTER 5
KANE'S EVENT-CAUSAL LIBERTARIANISM

The Elements of Kane's Theory

An agent having access to alternative possibilities is an intuitively plausible prerequisite for the agent's possessing free will. In general, when we consider ourselves as being able to do A at *t*, we think this entails the freedom to do other than A at *t*. For example, if I am free to check out a particular book from the library (A), this implies that I have the option ($\neg A$) of either checking out a different book or not checking out a book at all.

The most widely discussed libertarian theory in recent years is the event-causal account of Robert Kane. As an incompatibilist, Kane regards the ultimate responsibility of an agent as incompatible with determinism. However, he differs from many other incompatibilist proponents of free will in de-emphasizing the significance of alternative possibilities. Instead, Kane stresses a condition that Pereboom has characterized as *source-incompatibilist*, that an agent is responsible for an event or state if the agent makes an indispensable contribution to the production of the event or state:

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) did or omitted 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* (sufficient condition, cause or motive) for X, then the agent must also be personally responsible for Y. (Kane 1996, p.72)

The R clause specifies that E would not have occurred without some voluntary action or omission of the agent, while the U clause requires that the agent be responsible as well for any sufficient condition of E. For example, if an agent is to be responsible for burning

down a building, then something the agent voluntarily did must have causally made a difference to the burning of the building. And if a sufficient condition for the building burning down is that a source of ignition was applied to the walls of the structure in multiple spots, then the agent must be responsible for that as well. While UR does not entail that all of one's actions satisfy the alternative possibilities condition, we shall see below that Kane does require that some actions do satisfy it.

When an agent acts, her action is produced by, among other things, the determinate constitution of the agent. This constitution includes features such as the agent's habits, desires, intellectual capabilities, character and so on. Because these characteristics must already be in place at the time the agent acts, they may appear to be impediments to the freedom of the action. For example, suppose the agent's established character at the time of the action is such that the agent recoils from listening to popular music. Given this presupposition, if the agent is given the opportunity to listen to popular music, she will decline to do so. But then it seems that the agent was in fact not free to listen to popular music. The consideration that at any given time an agent's constitution is already determined by prior events has been developed by Galen Strawson into a general argument against the possibility of free will in the sense necessary for (ultimate) moral responsibility (Strawson 1986, pp. 27-28). It may be summarized as follows:

1. When an agent acts, the agent does so on the basis of her already established constitution.
2. But to be ultimately morally responsible, an agent must be ultimately responsible for her constitution.
3. To be ultimately responsible for her constitution, the agent would have to have consciously and explicitly chosen to have the constitution that she has.

4. But in order to consciously and explicitly choose to have the constitution she has, the agent would have to have consciously and explicitly chosen it by virtue of a previously established constitution, constitution -1.
5. And in order for constitution -1 to have been consciously and explicitly chosen, it must have been chosen by virtue of a still earlier constitution, constitution -2.
6. Clearly, an infinite regress ensues such that the agent cannot be ultimately responsible for her constitution, and in turn cannot be ultimately morally responsible.

Although the regress described is potentially infinite, finite agents such as human beings cannot embody an infinite regress. Eventually a point is reached when the agent does not yet exist. Hence the argument seems to show that a being that is not the ultimate cause of itself is not ultimately responsible for its constitution. Unless the regress can be halted during the lifetime of the agent, the kind of freedom necessary for ultimate moral responsibility does not appear to be possible. However, as we shall see, Kane's theory directly addresses this concern.

The central feature of Kane's event-causal libertarianism is the *self-forming action* (SFA). SFAs are

the undetermined, regress-stopping voluntary actions (or refrainings) in the life histories of agents that are required if U is to be satisfied, and for which the agent is personally responsible in the sense of R. The agents must therefore be responsible for them directly and not by virtue of being responsible for other, earlier actions (as would be required if they were not regress stopping). This means that, for SFAs, the "something the agents could have voluntarily done (or omitted) that would have made a difference in whether or not they occurred" is simply *doing otherwise*, rather than doing something *else* that would have causally contributed to their not occurring. (Kane 1996, p. 75)

SFAs are regress stopping because they emerge from indeterministic states of the agent. Kane argues that during times of indecision, the agent's cognitive processes at

the neural level are in a chaotic state that renders them sensitive to micro-indeterminacies.¹ These indeterminacies in turn screen off the will of the agent from deterministic influences (Kane 1996, pp. 129-131). According to Kane there are six kinds of SFAs in total, but the two most important are the ones that arise in cases of either moral or prudential conflict (Kane 1996, p. 125). As an example of an SFA arising from moral conflict, Kane describes a businesswoman who witnesses a crime taking place in an alley while she is rushing to an important meeting (*Four Views*: p. 26). Although she strongly wishes to attend the meeting, she has moral concerns regarding the crime victim. As a result she is torn between hurrying along to the meeting and calling for help. This moral conflict corresponds to an indeterminate state at the level of competing, parallel neural networks. The indeterminism screens off the influence of prior deterministic causation so that it is truly undetermined which of the two options businesswoman will select. In Kane's example, she ultimately decides to call for help. Although Kane by no means *proves* that the brain functions in the manner described, he does cite scientific evidence to support its plausibility (Kane 1996, pp. 129-130). Let us therefore grant the possibility of that the brain works in the manner Kane describes.

As we have seen, SFAs crucially involve indeterminate actions. One may doubt whether an agent can be responsible for an action if it is truly undetermined. On a natural interpretation of 'undetermined' it implies 'by chance' or 'by accident'; and if one's behavior arises by chance or by accident, then it is not sufficiently under the agent's control for the agent to be responsible for it. If, for example, a muscular spasm causes one's limb to suddenly jerk outward, one cannot be said to be responsible for it.

¹ We will return below to the question whether the regress-stopping characteristic of SFAs are sufficient to disarm Strawson's argument.

To this, Kane replies with the example of a man arguing with his wife who angrily slams his fist down onto a glass table with the intention of breaking it. We are to suppose that there is some indeterminacy in the motion of the arm which renders it uncertain up until the point of impact whether or not the glass will actually break. Suppose in the actual world the glass breaks, but in a nearby world the husband's counterpart fails to break it. Kane draws the following lesson.

Consider the husband and husband* (his counterpart in a nearby world who fails to break the wife's table). If the outcome is undetermined, husband and husband* also have "the same powers, capacities, states of mind, characters, dispositions, motives, and so on" up to the moment of breaking or not breaking the table, as the argument requires; and it is a matter of luck or chance that the table breaks in one world and not the other. But for all that, it does not follow, as (e) requires, that the husband is not responsible for breaking the table (Kane 1999, pp. 229-30). Although the arm motions of both husband and husband* are indeterminate they are the result of intentional, voluntary behavior. This distinguishes it from the case of an involuntary muscle spasm. Although in the case of husband, the breaking of the glass was undermined, insofar as it results from his voluntary, intentional act, he is responsible for its breakage. This shows that an agent can be responsible for his or her indeterminate actions.

Although SFA are the regress-stopping events that make ultimate responsibility possible, they are not the only actions for which we are ultimately responsible, according to Kane.

Luther's "Here I stand" would have been an affirmation for which he was ultimately responsible, even if it was determined and even if he could not have done otherwise, so long as it was a willed action (issuing from his character and motives) and he was responsible (as U requires) by earlier

undetermined SFAs for the character and motives from which the affirmation issued. In other words incompatibilists do not have to hold that all actions for which agents are ultimately responsible must be undetermined; ultimately responsible actions form a wider class of actions than undetermined SFAs. (Kane 1996, p. 77)

In addition to being regress-stopping, SFAs are also will-setting. That is, they establish the character and dispositions governing an agent's general, non-SFA conduct (Kane 1996, p. 78). Consider an agent who is tempted to peruse Facebook instead of performing her duty of writing a paper. This would be an example of prudential conflict. As in the case of moral conflict, the indecision is constituted by the chaotic disequilibrium of the agent's cognitive processing at the neural level so that it becomes sensitive to micro-indeterminacies. It is thus undetermined which option will be chosen. Suppose the option of working on her paper wins out. Apparently, on Kane's view this *sets* the agent's will so that in future circumstances when she is presented with a choice between working on her paper and perusing Facebook, she will be predisposed to work on her paper. Because her will was set by the SFA in question, the presentation of the same option in the future does not result in a state of indeterminacy. However, because the determinate pattern of response of the future instances was previously established by the regress-stopping of an SFA, the agent's later actions can be said to inherit ultimate responsibility from the earlier SFA. The agent is thus ultimately responsible for her later actions.

Objections to Kane's Theory

With the outlines of Kane's theory now in view, let's consider a version of the luck objection to event-causal libertarianism. A salesman has to decide whether to represent his product truthfully and possibly lose a sale, or misrepresent the product and increase his likelihood of making the sale. Suppose in the actual world the salesman decides, via

an indeterministic causal process, to misrepresent the product. Next, consider a nearby world at which the counterpart of the salesman, salesman* is confronted with the same choice, but decides instead to represent the product truthfully. Since indeterminism holds, the desires and reasons operative in salesman's decision making process do not determine what decision is ultimately reached. The same applies to salesman*. As we saw above, the mere fact that the choices were undetermined is no bar to their being responsible actions. Nevertheless, salesman and salesman* are identical to each other. Moreover, given the proximity of the possible worlds, both salesman and salesman* have the same desires with identical strengths and entertain the same reasons while possessing identical cognitive capabilities and tendencies. In light of these considerations, it seems to be a matter of luck that salesman chooses to misrepresent the product while salesman* decides to represent it truthfully. If this is so, then this example serves as an objection to Kane's theory. For in order for an action to count as a free action, it must be under the control of the agent. But if it is a matter of luck which alternative is selected from among a range of choices, then the action in question does not appear to be under the agent's control.

Kane has attempted to preempt an argument such as the above by arguing that given indeterminacy, exact similarity across possible worlds is undefined.

If the efforts are indeterminate, one cannot say the efforts had exactly the same strength, or that one was exactly greater or less great than the other. That is what indeterminacy amounts to. So one cannot say of two agents that they had exactly the same pasts and made exactly the same efforts and one got lucky while the other did not. Nor can one imagine the same agent in two possible worlds with exactly the same pasts making exactly the same effort and getting lucky in one world and not the other. (Kane 1996, pp. 171-2)

If Kane is correct, then the luck objection will be more difficult to formulate. However, contrary to Kane, Randolph Clarke argues that in physics, “the state of a particle with indeterminate position can be characterized by a state vector (or wave function), and a counterpart of the particle (or the particle in a different possible world) may be correctly characterized by exactly the same vector” (Clarke 2003, p. 86). Yet, even if Kane is right, the objection can be formulated in terms of the replaying of the event with the same agent and different outcomes, or simply by comparing agents who are indistinguishably similar to each other without being exactly the same (cf. Mele 1999, p. 280).

With this in mind, let it be understood that in the luck objection formulated above, salesman and salesman* inhabit distinct worlds that are indistinguishably similar to each other up until the divergence in the respective choices of the two agents whether to misrepresent or truthfully represent their products. Accordingly, salesman* is not strictly a different version of salesman, but is rather an agent who is so similar to salesman that the only difference we are able to observe is that salesman* makes a different choice than salesman when confronted with the option of being honest or lying. With this qualification in place it still appears that the only thing that accounts for the different choices of the two salesmen is luck. For there is nothing in the description of the either agent that explains why one speaks truthfully while the other misrepresents the product. And if luck is indeed what accounts for the choices of salesman and salesman* then it follows that they are not responsible for their decisions.

Kane responds to the luck objections by invoking dual efforts. He says that in moments of conflict the agent simultaneously tries to pursue both of the competing

alternatives. Although, it is undetermined which alternative will ultimately be chosen, the claim is that since each one is a voluntary effort of the agent, the agent is responsible for whichever one succeeds (Kane 1999, p. 231).

But these conditions are satisfied in the businesswoman's case as well, either way she chooses. If she succeeds in choosing to return to help the victim (or in choosing to go on to her meeting) (i) she will have "succeeded despite the probability or chance of failure"; (ii) she will have succeeded in doing what she was trying and wanting to do all along (she wanted both outcomes very much, but for different reasons, and was trying to make those reasons prevail in both cases); and (iii) when she succeeded (in choosing to return to help) her reaction was not "Oh dear, that was a mistake, an accident something that happened to me, not something I did." Rather, she endorsed the outcome as something she was trying and wanting to do all along; she recognized it as her resolution of the conflict in her will. (Kane 1999, p. 233)

Although this is an ingenious response, it is not altogether satisfactory. One doubt pertains to condition (iii). Whether or not the agent endorses the choice after it is made is an *effect* of the action, not a property of it. Thus it does not seem suitable as a criterion for the voluntariness of the action. But it is possible that Kane intends (iii) merely as a *symptom* of a voluntary action. If so, then he is right to draw attention to the fact that when one performs an action voluntarily, one will afterwards be disposed to endorse it. A more serious objection concerns the rationality of an agent actively willing two incompatible alternatives at the same time. Clarke argues

to have the agent actively making, at one time, two distinct efforts, one to (straightaway) make one decision and the other to (straightaway) make a contrary decision, is, it seems, to attribute to the agent a striking irrationality. It would be absurdly irrational for the agent to *make* the moral decision and, at the same time, *make* the self-interested decision, and it seems equally irrational for her to *try*, in distinct, simultaneous efforts, to (straightaway) make both. This added incoherence appears more of a threat than an aid to freedom. (Clarke 2003, p. 88)

Suppose I am trying to decide whether to order pasta or steak in a restaurant. Suppose I can only order one of the entrees due to having limited funds. Let us also assume that

I have an equal hankering for each. It seems that what I must do is decide *which* of the two I shall choose. Since I cannot have both, it would be irrational of me to attempt to order both dishes. Indeed, the very idea of trying to order both seems absurd. Instead, I will either try to find reasons for preferring one over the other, or else I will decide to make an arbitrary choice. But even when making an arbitrary choice, there is no question of trying to do both.

Against this, Kane might argue that simultaneously trying to decide between two incompatible options can be rational if the cognitive processing at work is not conscious. That is, suppose at the conscious level there is indecision about which choice to make, but that at a deeper, subconscious level of cognitive processing there are parallel efforts to select each option. If this were the case, then I might not be guilty of irrationality because the dual efforts would be below the threshold of conscious reflection. At the level of conscious awareness the parallel efforts would be manifested merely as indecision, and there is nothing irrational about having trouble making a decision. The problem with this response, however, is that it undermines my being responsible for the ultimate decision. SFAs are supposed to be exemplars of freely willed action. They are supposed to be the undetermined, regress-stopping actions that ground ultimate responsibility. But if they are non-conscious, it is difficult to see how they can play that role. In fact, if the cognitive processes that constitute the dual efforts are not conscious, they would appear not to be *my efforts* at all. On the contrary they would simply be unconscious mechanisms at work within me. Consciously, I would be in a state of indecision what to order. Then, with no further conscious participation on my part, one of the decisions would suddenly emerge victoriously. Given these considerations, it

seems Kane must admit that the efforts are conscious after all, but then, as per Clarke's objection, the dual efforts strategy will be found to be irrational.

Pereboom formulates another version of the luck objection in terms of a randomizing manipulator (Pereboom 2001, pp. 52-3). In this example, a version of Kane's businesswoman, businesswoman**, chooses to help the victim.² But businesswoman is under the control of a peculiar, randomizing manipulator. What the manipulator does is spin a dial that is integrated somehow with businesswoman**'s neural network. This dial's spinning is indeterministic. Moreover, its landing on a position is the neurophysiological realization of the choice to help the victim. Finally, "the dial's landing on a position does not supplement the agent's normal neurophysiological processes, but rather replaces the crucial indeterministic component in these processes" (Pereboom 2001, p 53).³

Arguably, conceiving of the indeterminacy in this manner on the one hand correctly characterizes the patterns of cognitive processing on the part of Kane's businesswoman and businesswoman**, and on the other hand highlights the alien character of the indeterministic feature of both of their decision making processes. When either businesswoman deliberates about whether to go to her meeting or to call for help, the indeterminism in her neural network seems qualitatively comparable to being manipulated by a randomizing device. In both cases, an indeterministic causal process that is not under the control of the agent is what brings about the ultimate

² Although businesswoman* is not mentioned, she would be a version of businesswoman who rushes ahead to the meeting without stopping to help.

³ Although Pereboom does not address the issue of the possibility of exact counterparts, as with the case of salesman and salesman* we can regard businesswoman** as being indistinguishable from businesswoman except for the former's neural network being integrated with the randomizing device.

decision. While the existence of the randomizing dial in the case of businesswoman** appears alien, peculiar, and not subject to determination by the agent, it helps to show that the natural indeterminacy in the case of businesswoman is also alien, peculiar, and not subject to determination by the agent. If the analogy between the two cases holds, the randomizing manipulator example will show that for both businesswoman and businesswoman** their respective decisions are not sufficiently under agential control to count as morally responsible actions.⁴

An immediate doubt about the analogy is that the randomizing manipulator, qua manipulator circumvents the will of the agent. Obviously, a libertarian should not be expected to regard a manipulated agent as free. But in this case, because the manipulation is limited to the randomizing of the choice, the manipulator does not exert any greater control over businesswoman** than Kane's businesswoman exerts over herself. In both cases, an indeterministic causal process selects which of two options the agent ultimately performs. In the case of the original businesswoman, her neural network is subject to micro-indeterminacies and thus it is undetermined which option will be selected. In the case of businesswoman** her neural network is subject to the indeterminacy of the spinning of the dial integrated with her neural network. Insofar as both the dial's spinning and the micro-indeterminacies in the neural network are indeterministic causal factors, they are on a par with each other. Intuitively, manipulation is thought to undermine an agent's freedom because it involves an external will circumventing the will of the agent. However, in the randomizing manipulator case, an

⁴ Another feature of this example is that it appears immune to the dual efforts strategy, because, intuitively, the mere fact that there are parallel efforts to perform two incompatible actions seems irrelevant to the fact that a dial spun by an external agent decides what action will be performed.

external will plays no greater role in the production of businesswoman**'s choice than the indeterminacy of businesswoman's neural network does in the production of her choice. This is because the randomizing manipulator reduces to the indeterministic spinning of a dial. The manipulator's will plays no direct role in selecting between the two options. And insofar as it plays an indirect role, all it does is trigger an indeterminacy that is qualitatively equivalent to the indeterminacy involved in the production of the original businesswoman's choice.⁵

Another criticism of Kane's theory involves its utilization of partially random events. In *Four Views*, Pereboom writes:

Let us designate those events for which factors beyond the agent's control determine their occurrence alien-deterministic events, and those that are not caused by anything at all truly random events. And we might call the events in the range between these two extremes—those for which factors beyond the agent's control causally influence their production but do not causally determine them, while there is nothing that supplements the contribution of these factors to produce them—partially random events. By the source-incompatibilist intuition, an agent cannot be morally responsible for a decision if it is an event that lies anywhere on this continuum, because the agent does not have a suitable role in its production—the agent will not be the source of such a decision in a sense sufficient for moral responsibility. But it seems for Kane free decisions are in fact partially random events. (*Four Views*: 108).

Kane's SFAs are partially random events because they essentially involve indeterminism with respect to which of two options is selected. When an agent reaches a state of indecision, this higher order fact maps onto chaotic turbulence in the agent's neural network. This chaotic turbulence is what renders the neural network sensitive to micro-indeterminacies. These indeterminacies are resolved in the definitive selection of

⁵ A further concern about the randomizing manipulator example is that businesswoman** might not experience her decision as resulting from her own will, but as Pereboom rightly says, it can simply be stipulated that she does experience it as coming from her own will.

one of the options which gave rise to the indecision in the first place. Since this choice results from an indeterministic causal process, it is at least partially random. This partial randomness diminishes the control the agent has over the production of the decision, for the agent does not get to control *which* of the two options is selected. The agent considers both choices, but the indeterministic causal process is what produces the final choice.

Kane rejects the claim that agents do not play a further causal role in bringing about their decisions over and above their psychological states:

For in *making efforts* of will to choose in terms of their reasons and motives, agents *do* play a causal role in bringing about their choices over and above the causal role played by their reasons, motives, intentions, and other mental *states* alone...Efforts are different from *desires* and other motives in this respect, contrary to what Pereboom suggests, because efforts are actions of the agents and not merely *states*. (*Four Views*: 174)

The contrast between efforts and desires is crucial to Kane's defense. Desires are passive states that an agent has, while efforts are active actions by the agent. One may have a desire and not act on it, whereas an effort is something one is actively pursuing. Pereboom's rejoinder to this is that in spite of the differences between desires and efforts they are on a par with each other in the role they play in Kane's account (*Four Views*: p. 194). Strictly speaking, both are states of the agent, or else agent-involving events. Moreover, whether one has conflicting desires or makes conflicting, dual efforts, if the factor that settles which desire or effort wins out is not under the control of the agent, then the agent is not morally responsible for it. This is the lesson of the randomizing manipulator example. In spite of the presence of parallel efforts, the indeterminacy that produces the decision is as alien, peculiar, and not subject to determination by the agent as would be a randomizing dial.

An interesting further objection raised by Mele is that Kane neglects to specify what a free action is, and that that he would have difficulty in doing so.

Kane does not claim that in cases of dual efforts to choose, the choices made are the products of freely made efforts. Nor has he put himself in a position to claim this, for he has not offered an account of what it is for an effort to choose to *A* to be freely made. Thus there is a salient disanalogy between cases like that of Kane's assassin and Kane's dual trying cases: there is no presumption that the dual efforts to choose are freely made. And if the agent's efforts to choose in a dual trying scenario—unlike the assassin's effort to kill the prime minister—are not freely made, it is hard to see why the choice in which such an effort culminates should be deemed free. (Mele 2006, p. 52)

In his reply, Kane appeals to compatibilist conditions. He argues that a free action can be viewed as one that satisfies a good compatibilist set of sufficient conditions for free action. These conditions, also known as “plurality conditions” include the ability to do whichever option they choose (or will) to do, when they choose to, for the reasons they choose, on purpose rather than by accident, mistake, or chance, without being coerced or compelled either in doing or willing the option, or otherwise controlled in doing or willing it by any other agent or mechanism (Kane 1996, p. 143). Inasmuch as Kane is a libertarian, he may appear surprising that he relies on compatibilist conditions to specify what a freely willed action is. As one might expect, critics have argued that an *incompatibilist* theory should not be so dependent on *compatibilist* conditions. Following a suggestion by Clarke (2003, pp. 89-90), Pereboom puts forward what Kane has designated “the-no-further-power” objection: “if decisions were indeterministic events, then agents would have *no more control* over their actions than they would if determinism were true, and such control is insufficient for responsibility” (*Four Views*: p. 107). Obviously, this conclusion cannot be acceptable to an incompatibilist.

Against Clarke and Pereboom, Kane contends that the power exercised under plural voluntary control is greater than what would exist in a determined world.

plural voluntary power that may be exercised in acting *or* acting otherwise, given the laws and the past of the *actual* world at the moment of action, *is* further power than the merely *hypothetical* power to do otherwise that compatibilists can give us in a determined world. Compatibilist power to have done otherwise may have been exercised only if the past or the laws *had been different in some way*, and agents do not have the further power at the moment of action to change the actual past or the laws of nature by their present actions. (*Four Views*: 177)

Moreover, Kane argues that the reason determinism precludes the kind of freedom needed for moral responsibility is because it rules out alternative possibilities. While Kane's theory does not give alternative possibilities as central a role as in some other theories, they are required for SFAs (Kane 1996, p. 72).⁶ The no-further-power-objection will thus have to be regarded as indecisive against Kane.

In addition to the preceding objections, I have formulated what I term *the first effort* objection against Kane's theory.⁷ This argument says in effect that while later SFAs are said to derive their freedom from earlier SFAs, there is no way to ground the freedom of the first supposedly free action. When an agent performs her first SFA, the indeterminism of her state of indecision merely renders it undetermined which of two options will be selected. The character of the agent thus far established will be determined by prior conditions. Likewise, the intensity of effort exerted with respect to each alternative will be determined by previous conditions. In addition, while the

⁶ Although an alternative possibilities condition for moral responsibility may seem intuitively plausible, ever since Frankfurt's seminal essay "Free Will and the Concept of a Person," it has been subject to attack. Because of the extensiveness of the Frankfurt argument literature, an evaluation of the bearing on Frankfurt-type argument's on Kane's theory will have to be left to another occasion.

⁷ See Clarke 2003, pp. 89-91; Pereboom 2001, pp. 49-50 and 2008, pp. 108-110 for similar objections to Kane.

indeterminacy in the agent's neural networks renders as undetermined which choice will get selected, the fact that the agent's choice bifurcates into just the options that it does is also determined by prior events. Thus there is really no freedom of the will being exercised in the first SFA. Not the character, nor the degrees of effort, nor the menu of options is undetermined. The only undetermined element is *which* of the two options gets selected. And that too is not under the control of the agent. Recalling Strawson's argument, it can be said that in numerous respects, the agent is not ultimately responsible for her constitution, for with respect to many aspects of her constitution, she has never had the opportunity (as this is her first SFA) to have consciously and explicitly chosen to have the constitution that she has. And if she is not free at the time of her first SFA, she is not free in any later SFA as there is no freedom for the later SFAs to inherit.

Kane's reply invokes the gradual emergence of responsibility starting with the limited level of responsibility possessed by children.

Thus, responsibility accumulates in human beings as they get older and build up a backlog of self-formed character. The only exceptions to this twofold source of responsibility are the earliest SFAs of childhood in which a backlog of *self-formed* character does not exist. In these earliest SFAs, all the responsibility is thus in the effort itself and the endorsement of its outcome of the agent when it succeeds. But precisely for this reason the responsibility for these earliest SFAs of childhood is not as great as later ones; in fact, in the earliest SFAs responsibility is minimal. (*Four Views*: 174-5)

Kane conceives of ultimate responsibility as having a small beginning and then gradually accumulating. But the defect in this conception is that there is no ultimate responsibility at all created at any stage. The will-setting that takes place in the first SFA, as well as in all subsequent SFAs, is a combination of determined character plus an indeterminate causal factor that is not under the agent's control. Although the will is progressively set beginning with the first SFA, the will thus constituted is never free.

Arguably, with the increasing determinacy of the will due to being set by successive SFAs, the occasion for indecision is progressively decreased and with it the influence of indeterminacy also decreases. Admittedly, the indeterminacy of the state of indecision creates alternative possibilities. But the mere fact that a decision may go in one of two directions is insufficient to generate freedom. According to source incompatibilist intuitions, for a choice to be free the agent in question must be ultimately responsible for it. But the agent cannot be ultimately responsible for it if everything about it, including what the options are, is determined except for which of the two options will ultimately be selected.

Conclusion

In summary, we laid out the main details of Kane's event-causal libertarianism. We put forth five arguments against it: Strawson's impossibility argument, the luck objection, the randomizing manipulator example, the no-further-power objection, and the first effort objection. We examined Kane's attempt to preempt the luck objection, but found ways around that attempt. We also considered Kane's dual efforts strategy and concluded that that reasonably persuasive considerations could be leveled against it. Also, the randomizing manipulator example seemed to be immune to the dual efforts defense. With respect to the no-further-power-objection, we conceded that Kane's defense against it is intuitively plausible, though a final assessment requires a thorough consideration of literature on Frankfurt-type arguments. Finally, we presented the first effort objection. This objection borrows some of its force from Galen Strawson's impossibility argument. This objection, along with Strawson's, appears to be decisive against Kane.

CHAPTER 6 POSTSCRIPT

The overarching purpose of my dissertation is to provide an account of the major incompatibilist free will positions that are available. I showed that these positions face serious and in some cases even insurmountable problems. As regards the position of compatibilism (in which we are able to maintain our concept of freedom of action while still adhering to determinism) while I covered only some of the debate about that view in my first chapter, my aim is more modest. Given the constraints of time and length, I cannot claim to entirely set aside this position nor show that the position can be successfully blocked. But I do think it was important to show the reader why the compatibilist position faces such serious challenges that one might, in seeing the nature of these challenges, be motivated to explore the incompatibilist route with regard to the debate. In the subsequent chapters, incompatibilist, or libertarian, theories of free will were refuted one by one. This leaves us to at least tentatively adopt the no-free-will thesis. But what does it mean to deny that we have free will; what are the implications of such a position? At present I will devote a few words to contemplate what life without free will would be like.¹

The Treatment of Criminals

If free will does not exist, then for most theorists, moral responsibility would be ruled out as well.² If this is the case, one question that arises is how criminals are to be dealt with. Ordinarily, sanctions against criminals presuppose that they are truly

¹ The present chapter is greatly indebted to Pereboom's discussions in the final two chapters of *Living Without Free Will*.

² John Martin Fischer stands out as the proponent of *semi-compatibilism*, the view that regardless of whether free will and determinism are compatible, moral responsibility and determinism are.

deserving of them. But in the acknowledged absence of free will, it may seem that genuine desert is ruled out. And if desert is ruled out, perhaps punishment would be as well insofar as intuitively punishment presupposes that agents are responsible for the actions they are being punished for.

Following Pereboom (2001, pp. 158-186) I argue that a solution to this puzzle is that criminals can still be held to be responsible for their actions in a causal sense. For example, in a case of shoplifting, whoever caused the item to be smuggled out of the store without paying for it is causally responsible for its being shoplifted. This principle of causal responsibility in turn provides a ground for sanctions against the shoplifter for the purpose of (a) discouraging her from doing it again, and (b) discouraging others from following her example. Punishment on this basis does not require us to attribute desert in a deep sense to the agent. All that is required is that the agent can be appropriately connected to the deed and for there to be sufficient social interest in discouraging the behavior in question. By punishing the causally responsible agent, a motivation is supplied for her to avoid such behavior in the future. Similarly, when others note the consequences of such behavior they will also be motivated to avoid it.

An additional justification for sanctions in the acknowledged absence of free will is based on an analogy with medical quarantine. It is well established that carriers of contagious diseases can be quarantined for the purpose of public safety. Similarly, those who exhibit a proclivity for criminal behavior can be quarantined in order to protect the public. In the case of medical quarantine, it is unnecessary to *blame* the contagious individual for being a carrier of the disease. All that is needed is to recognize the agent's causal role. Similarly, in the case of a habitual criminal, it would not be necessary to

blame the criminal in a sense that would presuppose free will. Instead all that is needed is to recognize the criminal's persistent causal roles in the commission of crimes. Although quarantine will resemble punishment in that it would restrict the agent's freedom, it would not strictly be punishment and would not be as harsh as true punishment would be.

A third principle for dealing with criminals in the acknowledged absence of free will is that of rehabilitation. This is similar to punishment in that it helps the criminal acquire a motivation to refrain from criminal activity. But it differs in that it involves explicitly applying behavior modification techniques to the agent in order to eliminate the undesirable behavioral trait(s). Using such methods, agents can be trained on the one hand to resist undesirable impulses and on the other to acquire beneficial habits. Obviously, such behavior modification techniques should be employed only if they are both humane and effective.

Attitude Towards Life

In addition to questions of how to deal with criminals, the acknowledged absence of free will would bring questions about the meaning of life. Although he is a hard determinist, Honderich recounts some of the challenges posed by such a view.

We have a kind of life-hope which is incompatible with a belief in determinism. An open future, a future we can make for ourselves, is one of which determinism isn't true.

Suppose you become convinced of the truth of our theory of determinism. Becoming really convinced will not be easy, for several reasons. But try now to imagine a day when you do come to believe determinism fully. What would the upshot be? It would almost certainly be dismay. Your response to determinism in connection with the hope would be dismay. If you really were persuaded of determinism, the hope would collapse.

This is so because such a hope has a necessary part or condition on which the rest of it depends. This is the image of origination. There can be

no such hope if all the future is just effects of effects. It is for this reason, I think, that many people have found determinism to be a black thing. John Stuart Mill felt it as an incubus, and, to speak for myself, it has certainly got me down in the past. (Honderich 2002, p.94)

I agree with Pereboom (2001, pp. 188-196) that there are reasons for thinking that acknowledging the absence of free will would not have such dire consequences. One consideration is that a great deal of what we are and what we experience would not be the results of free will even if free will were a reality. Our talents, the conditions of our upbringing, and the constraints posed by society are all factors that would be independent of free will even if we had it. Free will, if it existed, would only punctuate or lives at occasional intervals. By far, the major extent of our lives would still be determined. So, one reason for welcoming the absence of free will is recognizing how limited a role it would have played. For example, in choosing a career, whether or not I have free will, my talents and desires are held constant. Likewise, what employment opportunities are available would also be the same given either possibility. Hence what job I actually get would be the same whether or not free will exists.

In the acknowledged absence of free will, our future lives would still be a conjunction of our essential nature and the configuration of the world. In fact, it may be possible to take solace in determinism, as the ancient Stoics did. If there is no-free-will, then everything that happens may be attributed to either God or the universe itself. By identifying with this entity a profound equanimity may be attainable. Applying this to the career choosing scenario, since whatever job I actually get would be the same whether or not I had free will, in the acknowledged absence of free will I would be at peace with the outcome in virtue of seeing it either as the will of God or the unfolding of the universe—provided I identified with said entity.

Inter-Personal Relationships

An additional concern about acknowledging the absence of free will is that it may lead to our adopting an overly objectifying attitude towards other. In this regard, Peter Strawson writes:

What I want to contrast is the attitude (or range of attitudes) of involvement or participation in a human relationship, on the one hand, and what might be called the objective attitude (or range of attitudes) to another human being, on the other. Even in the same situation, I must add, they are not altogether exclusive of each other; but they are, profoundly, opposed to each other. To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided, though this gerundive is not peculiar to cases of objectivity of attitude. The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. If your attitude towards someone is wholly objective, then though you may light him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him. (Strawson 1962, p. 5)

Such an objective attitude is clearly dehumanizing. Thus it would be a tragedy if it were the inevitable result of acknowledging the absence of free will.

Fortunately, it is unlikely that human beings would ever renounce the subjective point of view in interacting with others. Strawson goes on to say:

And our question reduces to this: could, or should, the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way? For this is the only condition worth considering under which the acceptance of determinism could lead to the decay or repudiation of participant reactive attitudes.

It does not seem to be self-contradictory to suppose that this might happen. So I suppose we must say that it is not absolutely inconceivable

that it should happen. But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. (Strawson 1962, p. 7)

Along with Pereboom (2001, pp. 199-203), I think Strawson is largely correct in stating that the range of reactive attitudes and feelings necessary to preserve inter-personal relationships would persist even if the absence of free will were acknowledged. These reactive attitudes and feelings are simply too ingrained in human nature to disappear merely because of a change in philosophical belief.

While I agree that the reactive attitudes and feelings would not disappear, they may undergo some degree of evolution. For example, the attitude of taking offense or feeling indignation when mistreated would not disappear altogether. But they may undergo change so as to express more of a sense of disappointment and irritation than *pure* anger at the offending party. We may reach a point where we no longer *blame* the other person so much as experience a wish that the other treat one better combined with an ineradicable feeling of irritation at the other's behavior. Nevertheless, as Strawson suggests, one could still be said to be angry with the other person. Likewise, even with the acknowledged absence of free will people could still quarrel with one another. The quality of the quarrel may change slightly, because we would recognize that the other person is determined to believe as they do. But the absence of free will does not undermine our capacity to be responsive to reasons, so not knowing what reasons the other may accept, one would continue to vigorously press one's case. Moreover, human nature is such that residual irritation would be retained if the other

party failed to become convinced, so the affect of the quarrel would persist. Similar considerations apply in the case of forgiveness. Although one cannot strictly forgive where one does not blame another, the emotional character of the transaction would be retained. One would overcome an initial state of resentment for the other person's act, and transmute that into an attitude of renewed acceptance. In addition, the function of forgiveness in that one does not retain a grudge would continue. In other words, one would no longer seek sanctions against the other party for that which they have been forgiven. Or consider the attitude of gratitude. This would be retained, albeit in slightly changed form, even upon the acknowledged absence of free will. On the one hand, a person's gracious act would be seen as flowing from a pre-determined nature. On the other hand, the person will still be held to be causally responsible for the gracious act and a sense of pleasure and thankfulness for being the recipient of such graciousness would persist. Hence gratitude would retain most of its distinguishing characteristics. Similar considerations apply to all of the reactive attitudes and feelings that undergird inter-personal relationships. They would persist even if we came to accept that we do not have free will.

How Things Would Change

While I am largely in agreement with Pereboom, I would like to emphasize, to a greater extent than he does, the ways in which things would be different if we came to accept that we don't have free will. Pereboom writes:

First, as I have been arguing, hard incompatibilism leaves intact much of what in human life is especially important to us – morality, preventing human evil, the possibility of meaning and fulfillment in life through achievement – and there is more to come. If indeed the important things in life are not undermined by the belief that hard incompatibilism is true, this sort of pragmatic justification will be defeated. (Pereboom 2001, p. 198)

He is arguing here against proposals made by Kant, Smilansky, and Mele that a pragmatic justification for free will would override any metaphysical commitment to the no-free-will thesis. Pereboom's reason for rejecting the pragmatic arguments is that "much of what in human life is especially important to us" would remain intact. While I concede that the pragmatic arguments do not hold sway, I think Pereboom underestimates the existential implications of the no-free-will thesis. The genuine acceptance of the no-free-will thesis will profoundly affect our self-conception, as well as the way in which we view our daily lives.

Our not having free will means that everything we do is the product of what went before. Thus there is a qualitative similarity between how our lives are under determinism and how they would be if we were predestined. *Destiny* implies that what happens is prescribed by God or some other force. Thus there is nothing anyone can do about one's destiny. Also, according to one way of conceiving of destiny, a certain event that is predestined must occur, no matter what happens. The thesis of determinism does not imply this, because something different will happen if the causes are different. One cannot say that an event will occur no matter what. However, it could be argued that on a sensible conception of destiny, the causes themselves are predestined, and that is therefore through the action of those causes that the event in question takes place. This way of looking at destiny does cohere with determinism. Moreover, it yields the conclusion that whatever does in fact take place had to take place in virtue of what went before. Consequently, everything that happens in my life had to happen. This is a remarkable way to look at things. It also means that there is nothing I could have done to change any outcome for the reason that I was unable to change any of the causes.

This way of looking at things is profoundly different from how things seem when it is believed that I do have free will. On this latter conception, a great deal seems up for grabs, flexible and alterable. But on the no-free-will thesis, everything is given, predestined, for the past and the configuration of the world is already given at any given moment. Although Pereboom might concur with what I have said, he doesn't highlight these considerations.

Another thing that would be profoundly impacted if the no-free-will thesis is accepted is the status of deliberation. On the assumption that we do have free will, the outcome of deliberation is open. Either of two alternatives is truly possible. But given the no-free-will thesis, the outcome of the deliberative process is predetermined.

Admittedly, even in the absence of free will, I do not know what the outcome would be. So whether or not free will exists, the *results* of deliberating would be epistemically open; they would be unknown to the deliberator while deliberation is taking place. Nevertheless, in the absence of free will there is a sense in which I would be in error in thinking that what is epistemically open really is so. This forces me to conceive of deliberation in a new way.

Related to this, and perhaps even more significant existentially, is the implication accepting the no free will thesis on the very conception of my will. Regardless of one's metaphysical conclusions, it is undeniable that it *seems* that we have free will. This sense of having free will is robust and persistent. Almost certainly, this sense will remain even after the no-free-will thesis is accepted. Therefore the sense of free will will have to be regarded as an illusion. In this regard it has to be admitted that the no-free-will-thesis is an error theory. It says that we are fundamentally in error about an immediate

facet of our consciousness. I think Pereboom does not emphasize this consideration sufficiently. Accepting the no-free-will thesis will require us to regard a central fact of our consciousness as an illusion.³

Conclusion

Although the main goal of this dissertation has been merely to show that we do not have free will, we have briefly surveyed some of the implications of not having it. We raised questions concerning how such a realization would affect the treatment of criminals, one's attitude towards life, and inter-personal relationships. We saw that while accepting the no-free-will thesis would have an impact on each of these areas, the respective impacts would be benign. Nevertheless, the recognition that we do not have free will would have profound implications and would compel us to conceive of ourselves and our everyday lives in a new way.

³ This may be contemplated as an objection to the no-free-will thesis, but there are many occasions when theoretical considerations trump immediate perceptions.

REFERENCE LIST

- Balaguer, M. (2004). A coherent, naturalistic, and plausible formulation of libertarian free will. *Noûs*, 38: 379–406.
- Bok, Hilary, (1998). *Freedom and responsibility*. Princeton: Princeton University Press.
- Broad, C. D. (1952). Determinism, indeterminism, and libertarianism. In *Ethics and the History of Philosophy* (pp.195-217) London: Routledge and Kegan Paul.
- Campbell, C.A. (1967). *In defense of free will*. London: Allen & Unwin.
- Clarke, R. (1996) Agent causation and event causation in the production of free action. In Derk Pereboom (Ed.) *Free Will* (pp. 273–300). Indianapolis: Hackett.
- Clarke, R. (2003). *Libertarian accounts of free will*. Oxford: Oxford University Press.
- Clarke, R. Incompatibilist (Nondeterministic) Theories of Free Will. In *The Stanford Encyclopedia of Philosophy*. E. Zalta (Ed.), <http://plato.stanford.edu/archives/fall2008/entries/incompatibilism-theories/> Accessed 10 Jan 2012
- Coffman, E. (2010). How (not) to attack the luck argument. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 13(2), 157-166.
- Coffman, E. & Warfield, T. (2005). Deliberation and metaphysical freedom. *Midwest Studies in Philosophy*, 29, 25-44.
- Coffman, E. & Warfield, T. (2007). Alfred Mele's metaphysical freedom? *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 10(2), 185-194.
- Copp, D. (1997). Defending the principle of alternate possibilities: Blameworthiness and moral responsibility. *Nous*, 31, 441-56.
- Copp, D. (2003). 'Ought' implies 'can', blameworthiness, and alternate possibilities. In D. Widerker & M. McKenna (Eds.), *Freedom, Responsibility, and Agency* (pp. 265-299). Aldershot: Ashgate Press.
- Crisp, T. M., and T. A. Warfield. (2000) The irrelevance of indeterministic counterexamples to Principle Beta. *Philosophy and Phenomenological Research* 61(1): 173-84.
- Dennett, D. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Montgomery: Bradford Books.
- Double, R. (1991). *The non-reality of free will*. New York: Oxford University Press.
- Double, R. (1996). *Metaphilosophy and free will*. Oxford: Oxford University Press.

- Ekstrom, L.W. (2000). *Free will: A philosophical study*. Boulder: Westview Press.
- Finch, A., and T. A. Warfield. (1998) The mind argument and libertarianism. *Mind* 107(427): 515-28.
- Fischer, J.M. (1994). *The metaphysics of free will*. Oxford: Blackwell.
- Fischer, J.M. (2006). Free will and moral responsibility. In D. Copp (Ed.), *The Oxford handbook of ethical theory*, Oxford University Press: Oxford.
- Fischer, J.M. (2011) The zygote argument remixed. *Analysis* 71 (2): 267-272.
- Fischer, J.M., Kane, R., Pereboom, D., & Vargas, M. (2007). *Four views on free will*. Malden, MA: Blackwell.
- Fischer, J.M. & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.
- Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66, 829-839.
- Frankfurt, H. (1971). Free will and the concept of a person. *Journal of Philosophy*, 68, 5-20.
- Ginet, C. (1962). Can the will be caused? *Philosophical Review*, 71, 49-55.
- Ginet, C. (1966). Might we have no choice? In K. Lehrer (Ed.), *Freedom and determinism* (pp. 87-104), New York: Random House.
- Ginet, C. (1989). Reasons explanations of action: An incompatibilist account. *Philosophical Perspectives*, 3, 17–46.
- Ginet, C. (1997). Freedom, responsibility, and agency. *Journal of Ethics*, 1, 85–98.
- Ginet, C. (2002). Reasons explanations of action: Causalist versus noncausalist accounts, In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 386-405), New York: Oxford University Press.
- Ginet, C. (2007). An Action Can Be Both Uncaused and Up to the Agent. In C. Lumer and S. Nannini (Eds.), *Intentionality, deliberation and autonomy: The action-theoretic basis of practical philosophy* (pp. 243-255). Aldershot: Ashgate.
- Ginet, C. (1990). *On action*. Cambridge: Cambridge University Press.
- Ginet, C. (1996). In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing. *Philosophical Perspectives*, 10, 403-17.
- Goetz, S. (1988). A noncausal theory of agency. *Philosophy and Phenomenological Research*, 49, 303–16.

- Goetz, S. (1997). Libertarian choice. *Faith and Philosophy*, 14, 195–211.
- Goetz, S. (2000). Naturalism and libertarian agency. In W Craig and J. Moreland (Eds.), *Naturalism: A critical analysis* (pp. 156-186), London: Routledge.
- Honderich, T. (1988). *A theory of determinism*. Oxford: Oxford University Press.
- Honderich, T. (2002) *How free are you?: The determinism problem*. Oxford : Oxford University Press.
- Hunt, D. (2000). Moral responsibility and unavoidable action. *Philosophical Studies*, 97, 195-227.
- Kane, R. (1996). *The significance of free will*. New York: Oxford University Press.
- Kane, Robert. (1999). Responsibility, luck, and chance: Reflections on free will and indeterminism. *Journal of Philosophy*, 96, 217–40.
- Kane, R. (2000). The dual regress of free will and the role of alternative possibilities. *Philosophical Perspectives*, 14, 57-80.
- Kane, R. (2002). Some neglected pathways in the free will labyrinth. In R. Kane (Ed.), *Oxford handbook on free will* (pp. 407-437). New York: Oxford University Press.
- Kapitan, T. (1986). Deliberation and the presumption of open alternatives. *The Philosophical Quarterly* 36, 230-51.
- Kearns, S. (2011) Aborting the zygote argument. *Philosophical Studies*.
<http://dx.doi.org/10.1007/s11098-011-9724-3> 08 April 2011. 08 April 2011.
 Accessed 17 May 2012.
- Kim J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives* 3, 77-108.
- Kim, J. (1993). *Supervenience and mind: Selected philosophical essays*. Cambridge: Cambridge University Press.
- Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge: MIT Press/Bradford Books.
- Lewis, D. (1981/2003) Are we free to break the laws? In G. Watson (Ed.) *Free Will* (pp. 122-129), Oxford: Oxford University Press.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8, 529–66.
- Maurer S. and H. Gutermann (2007). Libertarian positions : A reconstruction. Resource Document. Goethe-Universität Institut für Philosophie <http://www.philosophie.uni->

frankfurt.de/lehrende_index/Homepage_Willaschek/forschung/tagungen/mele_fe
b_07/Gutermann_und_Maurer.pdf Accessed 20 Jan 2012.

- McCall, S. (1994). *A model of the universe*. New York: Clarendon Press.
- McCann, H.J. *The works of agency: On human action, will, and freedom*. Ithaca: Cornell University Press.
- McKay, T. J., and D. Johnson. (1996) A reconsideration of an argument against compatibilism. *Philosophical Topics* 24(2): 113-22.
- Mele, A. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Mele, A. (1995). *Autonomous agents: From self-control to autonomy*. New York: Oxford University Press.
- Mele, A. 2005. Libertarianism, luck, and control. *Pacific Philosophical Quarterly*, 86, 381–407.
- Mele, A. (2006). *Free will and luck*. New York: Oxford University Press.
- Mele, A. & D. Robb (1998). Rescuing Frankfurt-style cases. *Philosophical Review* 107 (1), 97-112.
- Nahmias, E., Morris, S., Nadelhoffer, T. and Turner, J. (2004). The phenomenology of free will. *Journal of Consciousness Studies*, 11(7–8), 162–79.
- Nelkin, D. (2004). The sense of freedom. In J. Campbell, M. O'Rourke, & D. Shier (Eds.) *Freedom and determinism* (pp. 105–34). Cambridge, MA: MIT Press.
- Nelkin, D. (2007). Good luck to libertarians: Reflections on Al Mele's free will and luck. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 10(2), 173-184.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Belknap Press.
- O'Connor, T. 2000. *Persons and causes: The metaphysics of free will*. New York: Oxford University Press.
- O'Connor, T. 2000. *Persons and causes: The metaphysics of free will*. New York: Oxford University Press.
- O'Connor, T. (2002). Libertarian views: Dualist and agent-causal theories. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 337–55). New York: Oxford University Press.
- O'Connor, T. (2003) Review of *Living without free will*. *Philosophical Quarterly* 53, 308-310.

- O'Connor, T. (2007). Is It All Just a Matter of Luck? *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 10(2), 157-161.
- Otsuka, M. (1998). Incompatibilism and the avoidability of blame. *Ethics*, 108(4), 685-701.
- Pereboom, D. (1995). Determinism *al dente*. *Nous* 29, 21-45.
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
- Pereboom, D. (2007). On Alfred Mele's free will and luck. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 10(2), 163-172.
- Pettit, P. (1989). Determinism with deliberation. *Analysis* 49, 42-44.
- Smilansky, S. (2000). *Free will and illusion*. Oxford: Clarendon.
- Strawson, G. (1986). *Freedom and belief*. Oxford: Oxford University Press.
- Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy* 48, 1-25.
- Stump, E. (2003). Moral responsibility without alternative possibilities. In M. McKenna & D. Widerker (Eds.), *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities*. Aldershot (pp.139-158). UK: Ashgate Press.
- van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press.
- van Inwagen, P. (1994). When the Will is Not Free. *Philosophical Studies*, 75, 95-113.
- van Inwagen, Peter. (2002) Free will remains a mystery. In R. Kane (Ed.) *The Oxford Handbook of Free Will*, 158-77. New York: Oxford University Press.
- Wegner, D. (2002). *The illusion of conscious will* Cambridge, MA: MIT Press.
- Widerker, D. (1995). Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *Philosophical Review* 104(2), 247-61.
- Yaffe, G. (2007) Libertarian accounts of free will. *Journal of Ethics: An International Philosophical Review* 11. 4, 485-497.

BIOGRAPHICAL SKETCH

Kirk W. Walker graduated *summa cum laude* from Herbert H. Lehman College in 1996. While at Lehman, he won the Jerome Weinstock Prize in Moral Philosophy. He earned his master's degree in philosophy in 1999 at the University of California, San Diego. He completed the degree requirements for a doctorate in philosophy at the University of Florida in 2012 and was awarded the Doctor of Philosophy degree in philosophy in the fall of that year.