



Hard Incompatibilist Challenges to Morality and Autonomy

by Sean Jason Stapleton

This thesis/dissertation document has been electronically approved by the following individuals:

Pereboom, Derk (Chairperson)

Sturgeon, Nicholas Lee (Minor Member)

Ginet, Carl Allen (Minor Member)

HARD INCOMPATIBILIST CHALLENGES TO MORALITY AND AUTONOMY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sean Jason Stapleton

August 2010

HARD INCOMPATIBILIST CHALLENGES TO MORALITY AND AUTONOMY

Sean Jason Stapleton, Ph. D.

Cornell University 2010

The following dissertation consists of three largely independent papers that are united in the general purpose of developing and extending the hard incompatibilist view against compatibilist and libertarian positions on free will. Hard incompatibilism is the view that the free will required for moral responsibility is incompatible with both the truth of causal determinism and its falsity. My position differs from those of some philosophers who hold similar views (such as Derk Pereboom and Saul Smilansky) in that I believe that this incompatibilism extends to important aspects of human life beyond the traditional issues of free will and moral responsibility. I do not attempt to independently establish the core claims of hard incompatibilism (by, for example, arguing for the inability of compatibilist and libertarian accounts to capture what's required for free, morally responsible action). Rather, I attempt to establish the correctness of hard incompatibilism in relatively less explored areas in the free will debate – in particular, some central deontic moral concepts and rational deliberation. I argue in chapter 1 that determinism is incompatible with moral obligation and, by extension, the moral wrongness of actions. In chapter 2 I provide reasons for thinking that it should not be psychologically possible for an agent to deliberate rationally while believing her deliberative process to be causally determined. In the final chapter I assume, in contrast with the first two chapters, that determinism is false, and I examine the implications of this assumption for the predicament of the rational deliberator. I argue that an indeterministic conception of one's own agency is

psychologically incompatible with the sense of control required for rational deliberation.

BIOGRAPHICAL SKETCH

Sean Jason Stapleton was born in Melbourne, Australia. From 1998 to 2002 he attended Monash University in Melbourne, where he completed a Bachelors degree, majoring in philosophy. At Monash he developed an interest in the philosophical areas of ethics and free will. He was accepted to Cornell's Sage school of philosophy in 2003, where he completed a Masters degree in 2007. He began his doctoral research at Cornell under the guidance of Dr. Carl Ginet, and completed the bulk of it under the guidance of Dr. Derk Pereboom.

Sean has been married to Lea since 2003.

ACKNOWLEDGEMENTS

Extra special thanks to my advisors: Carl Ginet, for his help with the background research and the first chapter, as well as his invaluable suggestions throughout the entire process; Derk Pereboom, for his endless reading of my drafts, his patience and his insightful suggestions, and most of all for his encouraging me to have faith in my ideas – the dissertation would not have taken the shape it has without his guidance. Thanks to the faculty, administrative staff, and graduate students of the Sage School of Philosophy who have helped me in different but equally important ways, particularly Erin Taylor, Nick Sturgeon, Patrick Mayer, Dick Miller, and Scott MacDonald.

Finally, I am especially grateful for the support of my wife and family.

TABLE OF CONTENTS

Biographical Sketch.....	iii
Acknowledgements	iv
Table of Contents	v

Chapter One: Why Actions Can't Be Morally Wrong if Determinism is

True	1
1.1 <i>Introduction</i>	1
1.2 <i>The Derivation</i>	1
1.3 <i>Understanding 'Ought' Implies 'Can'</i>	5
1.4 <i>An Objection to OIC</i>	11
1.5 <i>Obligations We Are Causally Determined to Satisfy</i>	16
1.6 <i>Moral Wrongness and Moral Obligation</i>	28
1.7 <i>Conclusion</i>	37
<i>References</i>	38

Chapter Two: How Belief in Determinism Threatens Rational

Deliberation	40
2.1 <i>Introduction</i>	40
2.2 <i>Motivating the Problem</i>	40
2.3 <i>Belief in the Availability of Multiple Possibilities for Action</i>	43
2.4 <i>Deliberation-Compatibilism</i>	44
2.5 <i>Interpreting the Problem</i>	46
2.6 <i>What Is Deliberation?</i>	47
2.7 <i>Motivation and Control</i>	51

2.8	<i>Causal Efficacy</i>	55
2.9	<i>Integration and Detachment</i>	62
2.10	<i>A Unified Account – Control, Motivation, Detachment, and Dissociation</i>	71
2.11	<i>Conclusion</i>	77
	<i>References</i>	78

Chapter Three: The Consequences of Indeterminism for Rational

Deliberation	80
3.1 <i>Introduction</i>	80
3.2 <i>Why a Deterministic Self-Conception is Problematic for Deliberation: Difference-Making Control and Ultimate Control</i>	81
3.3 <i>Requirement of Belief in Ultimate Control for Deliberation, and How Such Belief Can Be Compromised</i>	83
3.4 <i>The Basics of Kane’s Account of Deliberation</i>	95
3.5 <i>Antecedent-Determining Control and Ultimate Control</i>	97
3.6 <i>The Search for Ultimate Control on Kane’s Account</i>	100
3.6a <i>The Irrelevance of AD Control</i>	100
3.6b <i>Initiation of the Taoist Effort</i>	102
3.6c <i>The Taoist Effort and CSCs</i>	111
3.6d <i>Regaining Control</i>	115
3.6e <i>Other Indeterministic Efforts</i>	118
3.7 <i>Intention-Formation</i>	120
3.8 <i>Simple Breakdown</i>	123
3.9 <i>Conclusion</i>	127
<i>References</i>	130

CHAPTER 1

WHY ACTIONS CAN'T BE MORALLY WRONG IF DETERMINISM IS TRUE

1.1 Introduction

The primary purpose of chapter 1 is to address the question of whether causal determinism is compatible with the moral wrongness of actions. I will argue that determinism is incompatible with moral obligation, and that, due to the dependence relations between these two moral notions, it is also incompatible with the moral wrongness of actions. If determinism is incompatible with moral obligation, then it is also incompatible with any moral notions that entail obligation, including the moral rightness of actions that we are under moral obligation to perform. Whether determinism undermines the moral rightness of all actions is unclear, but this issue is beyond the scope of this chapter.

1.2 The Derivation

The following derivation is aimed at establishing these two claims. Premises (1) through (6) are intended to show that on the assumption that determinism is true, agents cannot be under moral obligations. The final conclusion (7) follows from (6) as an assertion of a dependence relation between moral obligation and moral wrongness.¹ The derivation is adapted from several sources.² It is as follows.

¹ I focus here on actions as opposed to some other standard locus of moral evaluation such as the character of the agent or the consequences of the action. I do not try to construct similar arguments for these other targets of moral judgment. Of course it is the case that, although moral rightness and wrongness attach to actions, we typically reserve judgments about goodness and badness, vice and virtue etc., for assessing character and consequences of actions, which will not concern me here.

² These include Copp (2003); Fischer (2003); Haji (1999); and Widerker (1991).

- (1) Suppose some agent, Jack, performed some action, *A*, say, stealing from the poor-box.
- (2) If Jack's doing *A* is morally wrong, then he ought to have refrained from doing *A*.
- (3) If Jack ought to have refrained from doing *A*, then he was able to refrain from doing *A*.
- (4) If determinism is true, then nobody is able to act differently from how they in fact act.
- (5) So, if determinism is true, then given that Jack did *A*, he was unable to refrain from doing *A*.
- (6) So, if determinism is true, then it's not the case that Jack ought to have refrained from doing *A*. (3-5)
- (7) Therefore, if determinism is true, then Jack's doing *A* is not morally wrong. (2,6)

Premise (3) is of course controversial; defending it is one of the main tasks of this chapter. Before doing so, however, let me make three brief remarks about other features of the argument. First, the argument as it stands is silent on the issue of the compatibility of determinism with moral obligations we are causally determined to satisfy. The argument assumes that Jack's action is one that he was under a moral obligation to refrain from performing, and further, that it is for his failing to satisfy his obligation to refrain from performing the action that the action is wrong of him. But we also perform actions that serve to satisfy obligations (keeping a promise is an example), and even if the performance of such an action is determined, such that the agent in question could not have done otherwise than perform the action, the argument does not entail that this inability to do otherwise undermines the obligation. I believe,

however, that it *is* the case that determinism undermines these kinds of obligations, and will provide some reasons for this below (section 1.5). Second, the argument is also silent on the issue of the compatibility of determinism with the moral rightness of actions. It seems that right actions fall into several categories: actions that we are under an obligation to perform and it would be wrong not to perform, e.g., keeping our promises, supporting our families; supererogatory actions, which are morally right to perform, but we are under no obligation to perform them, and it is thus not wrong not to perform them, e.g., giving a high percentage of our income to charity; and non-obligatory morally right actions that are not supererogatory, such as rescuing one of two people, both of whom need our help, but only one of whom we are able to help.³ The first type of right action is by definition dependent on obligation, so it will not require much argument to show that the moral rightness of these actions is undermined by determinism if the above argument is successful. Supererogatory actions are a much more difficult matter, as by definition these are actions that we are under no obligation to perform. So too are non-obligatory non-supererogatory morally right actions. It is an interesting question whether the rightness of these kinds of actions is also incompatible with determinism, although, as mentioned above, beyond the scope of this chapter.

Third, regarding premise (4), for the purposes of this chapter determinism will be understood as the thesis that ‘there is at any instant exactly one physically possible future’ (cf. van Inwagen 1983, 3). Exactly how to specify what this definition involves is not a simple matter, but I think it is safe to say that however determinism is understood, it implies the inevitability of our actions, in the sense that at any point in

³ I intend this distinction between obligatory right actions and non-obligatory, non-supererogatory right actions to roughly reflect the distinction Kant draws between perfect and imperfect duties (respectively). I use the distinction as a means of classification only, and do not intend for it to carry any theoretical implications. See Kant (1785/1981, Ak 421).

time prior to the performance of an action there existed conditions that causally necessitated that that action would be performed precisely when, and in the manner in which, it actually comes to be performed. If determinism is true, this inevitability is true of all actions and all events. Thus when we act we are constrained by all the relevant causal antecedents to act in precisely the ways we do act, and are never able to deviate from the performance of the actions that we are causally constrained to perform. If this is the case, then it follows on any construal of ‘ability’ other than mere logical possibility,⁴ that we are never able to do anything but what we in fact do.

In section 1.3, I will examine and refine the maxim ‘ought’ implies ‘can,’ of which premise (3) is an instance. I will indicate what I consider to be plausible interpretations of its ‘ought’ and ‘can’ components, and in section 1.4 I will defend the maxim against a counterexample. In section 1.5 I will attempt to show that if determinism is incompatible with moral obligations that we are causally determined to violate, then it is also incompatible with moral obligations we are causally determined to satisfy.⁵ Having established that determinism is incompatible with moral obligation, in section 1.6 I will argue that actions or omissions can be morally wrong only if

⁴ It might be thought that ‘*S* can do *A* at *t*’ is true so long as there is no logical contradiction involved in the proposition that *S A*’s at *t*. It seems quite clear, however, that the commonsense notion of ability that we employ when making claims about what someone can do at a particular time takes into account causal constraints applicable to *S* at *t*. To use one of David Lewis’ examples, even though on one reading of ‘can’ it is true that I can speak Finnish (the ‘facts about my larynx and nervous system are compossible with me [as opposed to an ape] speaking Finnish’), on the reading that is relevant to ability at a time (another set of facts that includes the fact that I never learned how to speak Finnish), I cannot speak Finnish. See Lewis (1986, 77). If determinism is true these causal constraints are pervasive, so that there is nothing that we can do at any particular time except what we are determined to do. This is a contentious matter, about which I have more to say below.

⁵ It is important for me to establish that determinism is incompatible with *all* moral obligations rather than merely obligations we are causally determined to violate, for two reasons. One is simply for the sake of completeness. The other is due to my claim that determinism is incompatible with the moral rightness of actions that we are under moral obligation to perform. If the argument only establishes that determinism is incompatible with moral obligations we are causally determined to violate, then this incompatibility would not extend to any actions considered to be morally right.

agents are under moral obligation with respect to them. I establish this claim by defending it against an objection.

1.3 Understanding ‘Ought’ Implies ‘Can’

The key premise of the argument is premise (3), which states that if Jack ought to have refrained from doing *A* then he was able to refrain from doing *A*, and is an instance of the principle that ‘ought’ implies ‘can’ (OIC). This principle can be considered an analog of the *Principle of Alternate Possibilities* (PAP), which states that ‘a person is morally responsible for what he has done only if he could have done otherwise’ (Frankfurt, 1969). I believe that both OIC and PAP are equally very plausible, and that alleged ‘Frankfurt-style’ counterexamples to PAP fail to establish that there can be situations in which a person *really* could not do *any* morally relevant thing other than what he did, and yet is morally responsible for doing it (I will not provide an argument for this here). Essentially, the reasoning in both cases is the same, in that PAP can be seen as a version of OIC applied to moral responsibility.

Before looking at objections to OIC it will be worthwhile to try to formulate the maxim more precisely, which will enable us to see more clearly what it says, whether it is true, and whether it can do the work in the argument that I am intending for it to do. I will look first at the ‘ought’ or obligation component, and then at the ‘can’ or ability component. Ishtiyaque Haji identifies three important distinctions that must be borne in mind when considering the maxim. The first is a distinction between moral and other kinds of obligation, such as legal, prudential, aesthetic, and what he refers to as the ‘plain’ or ‘overarching’ ought. It is a contentious matter whether overarching obligations exist, but if they do they would seem to arise in situations where we believe ourselves to have conflicting obligations of different kinds, e.g.,

moral and prudential. We may be considering each of these obligations, realizing that we cannot satisfy them both, and asking ourselves, “Given that I have a moral obligation to do *A* and a prudential obligation to do *B*, and given that I cannot do both *A* and *B*, what ought I to do?”⁶ While OIC may be applicable to these other kinds of obligation (for instance, maybe we don’t have legal obligations at particular times when we are unable to satisfy them) it is only moral obligations that need concern us here. A second distinction captures different uses of the term ‘ought’: sometimes we use the term to express what Haji calls a moral ideal or desideratum. We might, for instance, exclaim that “There ought not to be evil in the world,” or “There ought to be enough food for everyone to eat,” believing that there is little chance that these ideals can be realized. OIC clearly does not apply to this sense of ‘ought.’ Following W. D. Ross, Haji also draws a distinction between ‘absolute’ or ‘overall’ moral obligation, and ‘prima facie,’ or what might more accurately be referred to as ‘pro tanto’ (‘as far as it goes’) moral obligation. This distinction relates to situations in which we perceive ourselves to have conflicting *moral* obligations. For example, we may be faced with a decision whether to lie in order to prevent a co-worker from losing his job. The obligation not to lie and the obligation to help the co-worker are considered pro tanto moral obligations; the overall obligation, then, is the function of the stringency of these pro tanto obligations. Again, it is these overall obligations and not the pro tanto ones that are relevant when considering OIC (cf. Haji 2002, 14-15).⁷

As Haji points out, these pro tanto obligations are “real” as opposed to “on first glance” or “on the face of things” obligations. However, there is yet another

⁶ Assuming that we do sometimes have these overarching obligations, I am agnostic, as is Haji, as to whether they are moral obligations. If they are, I see no reason why OIC should not also apply to them, in the sense that we cannot have overarching (moral) obligations that we cannot satisfy.

⁷ Copp refers to the kind of obligation I am trying to characterize by employing these distinctions as ‘moral requirement.’ According to Copp (2003, 273), to be morally required to do a particular thing is to have no morally acceptable option but to do that thing.

distinction that can usefully be made, by contrasting two different kinds of *judgments* of obligation, which we might refer to as *prima facie* and *all things considered* judgments of obligation. A *prima facie* judgment of obligation in this sense might be thought of as a hasty judgment, or a judgment made in the absence of a full appreciation of all the relevant facts, or perhaps a general judgment that would ordinarily apply to cases of a similar kind to the one under consideration. In David Copp's terminology, to judge that someone has an all things considered moral obligation [is morally required] to do a thing entails judging that the person has no morally acceptable option but to do that thing.

Some examples should help to clarify this distinction between *prima facie* and all things considered judgments of moral obligation. For instance, we may initially perceive ourselves as having an obligation to call the police when we see someone being attacked in the street, until we realize the people involved are stuntmen acting in a movie. Or, of more relevance to OIC, we may judge that a person has an obligation to refrain from screaming blasphemous obscenities in church, until we realize that she suffers from Tourette's disorder, and is unable to refrain. Or we may judge that a firefighter has an obligation to rescue someone from a burning house, until we realize that he has just now become paralyzed. As a justification for this interpretation of the 'ought' in OIC, as well as a motivation for the truth of the maxim so interpreted, we may consider what Copp calls the 'fairness' argument. The idea is that just as it would seem unfair to demand or require of a person that she do something she is unable to do (to require someone to, for example, lift a car above her head), there seems to be a kind of unfairness involved in a person's having a moral obligation she is unable to satisfy. True, agent-requirements are not necessarily moral requirements; in the moral case there is no one doing any requiring or demanding, no one who could be blamed

for being unfair. Nevertheless, as Copp points out, if it's not the case that there is any (moral) unfairness in an agent's being morally required to do something she can't do, it's hard to see why there would be unfairness in one agent's requiring of another that she do something she can't do (Copp 2003, 271-2, 278).⁸

There is also a need to refine the 'can' component of OIC. Haji formulates stronger and weaker versions of OIC in this respect. For my purposes the weaker of these will suffice. It is as follows.

Agent, *S*, ought to do something *A*, only if [1] *S* has the opportunity to do *A*, [2] *S* is physically and psychologically able to do *A*, and [3] *A*'s accomplishment is not "strictly out of *S*'s control" (Haji 1999, 182; 2002, 23-4).⁹

The first condition captures the idea that an agent cannot have an obligation to do something unless she is in a position in which she can exercise her ability to do it. 'The concern is that not being relevantly situated or not having the cooperation of the environment would prevent one from exercising the pertinent abilities' (Haji 2002, 22). Haji endorses Kadri Vihvelin's characterization of what it is to have the opportunity to do something: '[Opportunity] include[s] all circumstantial factors, both those inside and outside the skin, which would enable someone to exercise a relevant ability. On this view, someone lacks the opportunity to do *x* just in case there is something that prevents or would prevent a successful exercise of the relevant ability, regardless of whether it's something external (locked door, chairs, lack of piano) or

⁸ Understanding OIC in terms of moral requirement also helps to block objections to it such as Fischer's. Fischer claims that because the goal of practical reasoning is to determine what we have sufficient reason to do, the 'ought' in OIC should be analyzed in terms of what the agent in question has sufficient reason to do. Understood in this way, the proposition that *S* ought to do *A* is compatible with *S* being unable to do *A*. See Fischer (2003, 247-8).

⁹ The stronger version contains two additional requirements. These are, first, that the agent 'has the relevant 'know-how' to do *A*,' and second, that the agent is 'not ignorant of germane facts (this is meant to capture the 'know-that' requirement)'.

something internal (broken legs, cramps, state of unconsciousness)' (2002, 22).

Concerning the second condition, that *S* be physically and psychologically able to do *A*, Haji emphasizes that it does not include a 'know-how' requirement, and so does not amount to what Vihvelin calls 'volitional ability.'¹⁰ Haji does not give much in the way of a positive explanation of what it is to be physically and psychologically able to do something while lacking the volitional ability to do it. So I'll assume he intends this condition to exempt *S* from having an obligation if *S* is, for instance, paralyzed, tied up, or suffering from convulsions (on the physical side) or in the grip of a psychotic episode or extreme phobia (on the psychological side), i.e., any physical or psychological impairment or impediment that would prevent *S* from doing *A* *excluding* *S*'s simply not knowing how to do *A*. The final condition, that *A* not be 'strictly out of *S*'s control' is explained as follows: 'if it were always the case that when an agent formed a here-and-now intention – that is, an intention to perform an action right away – it became highly improbable that she would act as she intended, then even if she did succeed in doing that thing, her accomplishment of it would be strictly out of her control.'

I think these refinements of the notion of ability as it relates to OIC are quite plausible and (perhaps with the exception of the final condition) uncontroversial, although they are also compatible with the assertion that agents have the ability to perform actions they are causally determined *not* to perform. However, the commonsense notion of ability captured by these conditions does not take determinism into account. It seems to me that when determinism is taken into account there is a strong intuitive sense in which one lacks the ability to do that which one is causally

¹⁰ For Vihvelin, a person has the 'volitional ability' to do *x* 'just in case it's true that there are some reasonably specifiable circumstances *C*... such that *if* she tried, in circumstances *C*, to do *x*, she would (probably) succeed.'

determined not to do. Therefore, another refinement suggests itself. It is the requirement of what Fischer (in a somewhat different context) refers to as ‘access control,’ and is spelled out in terms of possible worlds. It is the idea that in order to have the ability to perform some action, an agent must have *genuine access* to the possible world in which she performs it, and one has genuine access to a possible world only if it shares the same past and natural laws with the actual world (Fischer 2003, 249). Given the above definition (page 3) if determinism is true then there is only one possible future, that is, only one possible world that anyone has genuine access to. So, if determinism is true, then one does *not* have the ability to do what one is causally determined not to do, since the world in which one does what one is causally determined not to do is not genuinely accessible.¹¹

With these refinements in hand, I think a number of objections to OIC can be adequately addressed, and in a similar manner. For example, one such objection arises from consideration of moral dilemmas. A moral dilemma can be characterized as a situation in which an agent *S* has a moral obligation to perform action *A* and an equally compelling moral obligation to perform action *B*, but cannot perform both. The reply involves both the abovementioned point about interpreting the ‘ought’ in OIC in terms of *all things considered* moral requirement, and also the argument from fairness. For a moral dilemma to constitute a counterexample to OIC it must be the case that both *A* and *B* are morally required of *S*. Given the above characterization of a

¹¹ This last refinement is likely to be resisted by proponents of *conditional* analyses of ability. These essentially involve the idea that all that is required in order for it to be true that an agent has the ability to perform some particular action that s/he in fact did not perform (or is causally determined not to perform) is that the agent *would* have performed it *if* some factor had been different, for example, if s/he had wanted to perform the action, or if some event in the past or the laws of nature had been different. I will not bother trying to refute these claims here. Roderick Chisholm has provided what I consider to be a decisive refutation of conditional analyses. See Chisholm (1982, 26-7). More recently, some writers have attempted to improve upon earlier versions of conditional analyses by interpreting claims about ability to do otherwise as involving dispositions to act, and offering conditional analyses of dispositions that, they claim, are compatible with determinism, and immune from the kinds of counterexamples often employed against dispositional analyses.

moral requirement, the objector must hold then that *S* has no morally acceptable option but to perform *both A and B*. But this is not possible for *S* to do. It seems that the objector must deny the interpretation of OIC whereby the ‘ought’ refers to a moral requirement, or at least to deny that this interpretation is applicable in cases of moral dilemmas. But the argument from fairness seems clearly to apply here too. Just as it seems unfair to require of someone that she perform an action she cannot perform, it seems equally unfair to require a person to perform two actions such that she cannot perform both (See Copp 2003, 278-82).¹²

1.4 An Objection to OIC

Pete Graham (unpublished) proposes an interesting argument against OIC, based on a certain form of counterexample. The form of the counterexample is based, in turn, on a certain sort of obligation that people can have:

‘Suppose that Jones’s [A]-ing is morally permissible only because Smith’s [B]-ing is (or would be) morally impermissible. If OIC were true, then Smith’s being rendered incapable of refraining from [B]-ing in such a case should render Jones’s [A]-ing morally impermissible. But, if it is not intuitive that so incapacitating Smith has this effect on the permissibility of Jones’s [A]-ing, that is good intuitive grounds for thinking that OIC is false’ (Graham, 7).

To illustrate this, Graham gives the following example:

‘TRANSPLANT: A surgeon has ten patients, each of whom will die of organ failure if they don’t receive an organ transplant. The surgeon really wants to save her patients and is convinced (perhaps brainwashed) by Consequentialist arguments to the effect that it would be morally permissible to kill two people

¹² There are a number of other objections discussed by Copp that can be responded to similarly, by attending to the refinements to OIC given above, which I will not discuss here.

in order to save them. “Luckily” she notices that in another room of the hospital there are two innocent and unconscious tonsillectomy patients who are perfect organ matches for her patients. The hospital janitor, aware of the situation, searches in vain for a way to prevent the surgeon from killing the two. As events are unfolding very quickly, it turns out that the only means by which he can stop her from chopping up the two innocent patients is by shooting her with his pistol. He does so and thereby kills her.’

Graham makes two points about this example:

(1) It is morally permissible for the janitor to kill the surgeon,
and the best explanation of this is

(2) if the janitor had not killed the surgeon, then the surgeon would have
morally impermissibly killed two innocent people (Graham, 8).

He then alters the example to TRANSPLANT (COMPULSION) which is the same as TRANSPLANT except that the surgeon is psychologically compelled to kill the two in order to save the ten. Graham claims that in TRANSPLANT (COMPULSION), just as in TRANSPLANT, (1) is true, and the best explanation of (1)’s being true is that (2) is true. Now with TRANSPLANT (COMPULSION) he thinks he has a counterexample to OIC, because if (2) is true, and if

(3) if the janitor had not killed the surgeon, then the surgeon would have killed
the two innocent people and been unable to refrain from killing them
is also true, then it must be the case that

(4) if the janitor had not killed the surgeon, then the surgeon would both have
morally impermissibly killed the two innocent people and have been
unable to refrain from killing them (Graham, 9).

Hence, OIC is false, as the surgeon is unable to refrain from killing the two innocent patients, and yet is under a moral obligation to refrain from killing them. Graham then

goes on to discuss a number of alternative explanations for the truth of (1) other than (2) and rejects them as being unsatisfactory.

If Graham's argument is sound, i.e., if TRANSPLANT (COMPULSION) is indeed a counterexample to OIC, then my argument is in trouble. However, I think there is reason to question whether this is the case. Graham recognizes that the crucial step in his argument is the claim that the best explanation of the truth of (1), that it is morally permissible for the janitor to kill the surgeon, is (2), that if the janitor does not do so, the surgeon will morally impermissibly kill the two innocent patients. It seems to me that in TRANSPLANT, this would be a good, intuitively plausible candidate for a correct explanation of the moral permissibility of the janitor's killing the surgeon, although perhaps not a complete one (I have more to say about this below). However, it is much less clear that (2) is the best, or even a plausible, explanation of (1) in TRANSPLANT (COMPULSION). This is in part because it is not obvious that genuinely compelled actions can be morally impermissible. Due to the kinds of considerations raised above with respect to the argument from fairness, if the surgeon was genuinely compelled, then it's not clear that her killing the patients is impermissible, and hence not clear that (2) is even *true*, let alone a plausible explanation of (1).

So what does explain why the janitor is permitted to kill the surgeon in TRANSPLANT (COMPULSION)? I propose that a reasonable alternative explanation is that the surgeon's intended action *seriously threatens the lives of two innocent people*. The janitor's action is not only permissible, but I think (setting aside concerns about determinism), obligatory, because it is generally obligatory for a person to do what they can to help neutralize serious threats to innocent human life. So I contend that Graham's (2) should be replaced, in this case, by:

(2`) if the janitor does not kill/stop¹³ the surgeon then the janitor will have morally impermissibly failed to have neutralized a serious threat to the two innocent patients;

and this is what explains (1) in TRANSPLANT (COMPULSION). And as far as TRANSPLANT is concerned I think it is the conjunction of (2) and (2`) that explains (1).

Furthermore, (2`) is a more straightforward explanation that seems to apply to a wider range of cases than does (2). That is, it seems clear that (2) is only (prima facie) relevant in cases involving threats posed by persons. If, for instance, the threat to the patients were posed by a poisonous snake it seems obvious that the permissibility of the janitor's killing the snake would not derive from any moral issue concerning the potential behavior of the snake; it would derive purely from the fact that the snake would be posing a serious threat to the patients. It is only in cases where the threat is posed by a person that any moral issue concerning the source of the threat can become relevant. And, as we have seen, it is not implausible that these moral issues can only play the kind of justificatory role at issue here in a certain subset of these cases, e.g., where the person posing the threat is able to refrain from carrying it out.

It may be objected here that my (2`) is really just *part* of the explanation of why the janitor's action is permissible. It may be thought that we also need the claim that the surgeon's killing the innocent people would, in the circumstances, be impermissible, and so Graham's explanation of (1) stands as the best candidate. In

¹³ The fact that the surgeon is a person, having rights, and not some non-human threat such as a disease or an avalanche, means that proportionality requirements must be adhered to. It is not permissible for the janitor to kill the surgeon if there is open to him a less harmful way of neutralizing the threat.

TRANSPLANT this does seem reasonable, but if it is doubtful, as I contend, that the surgeon's intended act of killing the two patients in TRANSPLANT (COMPULSION) is impermissible, due to its being compelled, and if there is a suitable, unproblematic alternative explanation available, then there seems to be no reason not to exclude Graham's (2) from the explanation of this case entirely.

It may also be objected that whatever explains (1) in TRANSPLANT must be the same as that which explains (1) in TRANSPLANT (COMPULSION). Graham thinks this seems obvious (Graham, 11). I don't see why it should be obvious. The cases are very different due to the inability of the surgeon to refrain in the latter case, and it is no news that our moral intuitions and moral judgments tend to vary significantly in response to even seemingly minor alterations to cases. So without some reason why the explanation of (1) must be the same in both cases, I disagree with Graham that it obviously should be. Even if I were to concede the point that the explanation in both cases should be the same I would argue that it is more reasonable to think that the explanation of (1) that is common to both cases is (2') rather than (2); (2) simply becomes more salient in cases such as TRANSPLANT, where moral impermissibility on the part of the source of the threat is relevant. I conclude here that Graham's example should not persuade us to doubt that 'ought' does indeed imply 'can'.¹⁴ In the next section I address the issue of whether we can have moral obligations we are causally determined to satisfy.

¹⁴ I am very grateful to Carl Ginet for showing me some correspondence between himself, Pete Graham, and Michael McKenna regarding Graham's paper. Some of my suggestions were inspired by objections to the paper made by Ginet and McKenna. However, I do not intend my objections to be representative of theirs.

1.5 Obligations We Are Causally Determined to Satisfy

By this point I claim to have established that determinism is incompatible with moral obligations that an agent is causally determined to violate. Here, again, is the derivation intended to establish that result (from page 2):

- (1) Suppose some agent, Jack, performed some action, *A*, say, stealing from the poor-box.
- (2) If Jack's doing *A* is morally wrong, then he ought to have refrained from doing *A*.
- (3) If Jack ought to have refrained from doing *A*, then he was able to refrain from doing *A*.
- (4) If determinism is true, then nobody is able to act differently from how they in fact act.
- (5) So, if determinism is true, then given that Jack did *A*, he was unable to refrain from doing *A*.
- (6) So, if determinism is true, then it's not the case that Jack ought to have refrained from doing *A*. (3-5)
- (7) Therefore, if determinism is true, then doing *A* was not morally wrong of Jack. (2,6)

I have argued that if determinism is true then at time *t* when I believe myself to have a moral obligation, it is already determined whether I will honor it or not. If it is determined at *t* that I will, at some future time, perform some action that violates my obligation, then OIC tells us that I cannot be under such an obligation (the derivation to (6)). But what should we think about cases where it is determined at *t* that I will *not* violate the obligation? OIC tells us only that I cannot be under an obligation to perform some action that I cannot perform (or to refrain from performing some action

that I am unable to refrain from performing). According to OIC, we need to have the ability to do *A* in order to be under an obligation to do *A*, so it may be thought that if we're causally determined to do *A*, then we have the ability to do *A*, and so we can be under an obligation to do *A*, even though we are unable to refrain from doing *A*. That is, although determinism and OIC entail that we can be under no obligation to do *A* if we're causally determined not to, we can still properly be under an obligation to do *A* if we're causally determined to do *A*, because we still *can* do *A*. If this is correct we would be in a position whereby we can be under an obligation to do only that which we are causally determined to do, yet these obligations would be legitimate.

In order to resist such a claim, Haji offers basically two considerations in favor of a principle he refers to as CK, which states that if one ought to do *A*, then one can refrain from doing *A* (and if one ought not to do *A*, then one can do *A*). I shall focus on the second, which seems to be the stronger of the two. This consideration is based on Fred Feldman's theory of what it is for an act to be obligatory. The idea is that when one is faced with a moral choice there are several possible worlds accessible to one, i.e., possible worlds that would exist if the particular choice required to bring them about is made. 'Roughly, a possible world is accessible to a person at a time if and only if it is still possible, at that time, for the person to see to the occurrence of that world.' A world is 'bypassed,' and thus becomes inaccessible, once a person behaves in some way in which she does not behave in that world.

' Ks,t,p ' abbreviates 'there is a world accessible to *s* as of [time] *t* in which state of affairs *p* occurs' (*K* being a kind of possibility operator); ' Us,t,p ' abbreviates '*p* occurs in every possible world accessible to *s* at *t*' (*U* being a kind of necessity or unalterability operator).

According to Haji, Feldman's theory of moral obligation is as follows:

MO: A person s ought, as of t , to see to the occurrence of a state of affairs p if and only if p occurs in all the intrinsically best worlds accessible to s as of t .

If we let ' $O_{s,t,p}$ ' abbreviate ' p occurs in all the intrinsically best worlds accessible to s at t ,' then MO validates OIC, since $O_{s,t,p}$ entails $K_{s,t,p}$. Furthermore, on this theory, unalterability implies obligation, i.e., if MO is true then

UO: $U_{s,t,p}$ entails $O_{s,t,p}$

is also true (Haji 2002, 28-31).

In other words, if p is unavoidable for you (if p occurs in every possible world now accessible to you) then p occurs in all the *best* worlds accessible to you. And if (according to MO) you ought to do a thing only if that thing occurs in the best worlds accessible to you, then it's not the case that you ought to do something if that thing *doesn't* occur in any world accessible to you, as in that case it also doesn't occur in any of the best worlds accessible to you (there must be an accessible world in which that thing occurs in order for you to have an obligation to do it). This is essentially OIC.

A potential problem for UO is that if $U_{s,t,p}$ entails $O_{s,t,p}$ then, in combination with MO, it may turn out that it is obligatory for you to bring about some morally abhorrent state of affairs, say, a murder, if it's the case that the murder occurs in all the best worlds accessible to you (which would be true if it was unavoidable, i.e., if it occurs in *all* the worlds accessible to you).

Due to this, Feldman proposes a different obligation operator, O^* , which functions as follows: ' $O^*_{s,t,p}$ ' is true if and only if p does not occur in all the worlds accessible to s as of t , but p does occur in all the best of these worlds.' So, you are

obligated to do a thing only if that thing occurs in the best worlds accessible to you, *so long as* there are other worlds accessible to you in which you don't do that thing, i.e., so long as you can avoid doing that thing. This, of course, supports CK, which says that you ought to do a thing only if you can refrain from doing it.

Despite the technical details, I can't see that the part of this theory that supports CK is anything more than an *ad hoc* modification employed to avoid an unpalatable consequence. Haji claims that 'if the theory does capture the nature of moral obligation, then there is a powerful theory-fuelled incentive to endorse both [OIC] and its complement CK.' I don't see anything obviously wrong with the theory (MO), however, it doesn't seem to be the theory itself that supports the introduction of O^* , but rather the fact that the theory delivers unintuitive results in cases where an agent's alternatives for action are restricted to one.

For these reasons, I don't believe that Haji's defense of CK would be adequate to persuade an opponent that we cannot have obligations we are causally determined to satisfy. Therefore, I offer the following analysis. However, it is not intended to be decisive. I offer it partly for the sake of completeness and partly because of my intuitive belief that determinism is incompatible with all of deontic morality (and *a fortiori* moral obligation): the idea that we can be obligated to do something that we will inevitably do seems to me only slightly less intuitive than the idea that we can be obligated to do things we won't do. Although the above derivation would be unaffected if it turned out that we can have such obligations, it is in this sense incomplete. It establishes only that determinism is incompatible with *some* obligations and not others, and so leaves it open that, if we can have obligations we are causally determined to satisfy, then the division between the perceived obligations we actually have and the ones we don't depends on actual (future) occurrences, i.e., whether or not

we turn out to perform, or not perform, the relevant actions that would amount to satisfying or violating them. Therefore, I offer these suggestions in an attempt to substantiate, to some extent, my belief that determinism is incompatible with all moral obligations.

If the derivation to (6) is sound, i.e., if we cannot be under an obligation to do something we are causally determined not to do, then accepting that we can be under an obligation to do something we are causally determined to do would leave us in a strange position. It would mean that if we are uncertain that we will do *A* and also uncertain that determinism is false (perhaps the position that many of us are in) then we should be uncertain that we have an obligation to do *A*. From a practical standpoint, our level of confidence in the possibility that we have the obligation would be so low as to render it incapable of playing any significant role in deliberation. The function of obligations is to guide actions, and so they must be able to count as reasons for deciding how to act, and must thus be able to be recognized by us as fairly reliably true (or false) before the time we make our decision about whether to do *A* (and thereby come to know that we will do *A*). But if we can never be sure whether we have an obligation, due to the fact that we can never know beforehand whether or not we are causally determined to satisfy it, then obligations would, for those uncertain that determinism is false, seem to lose their action-guiding usefulness. For example, suppose an acquaintance of mine, Jim, is sick and in the hospital, and I am wondering whether I have a moral obligation to visit him. As I am reflecting on the truth of determinism and OIC (let's say I am naturally somewhat unsympathetic), it may occur to me that if I am causally determined to visit Jim then I have an obligation to do so, but if I am causally determined not to do so then I have no such obligation, because I can't have an obligation to do something that I won't (can't) do. This leaves me in an

odd position. Since I have no way of knowing whether or not I am causally determined to visit him, the notion of moral obligation gives me no help in deciding upon the right thing to do.

It may be objected here that my requirement of certainty is unrealistically stringent. That is, perhaps we can, and often do, decide upon a particular course of action motivated solely by the belief that we *probably* have, or *may* have, an obligation to so act. If this were the case then it would be false that an agent's uncertainty as to the truth of determinism, combined with her acceptance of the derivation, would result in the undermining of the action-guiding usefulness of moral obligations she is causally determined to satisfy.

To illustrate this, consider the case of Sue the student. While doing her grocery shopping, Sue realizes that the cashier has accidentally given her an extra \$100 in her change. In the process of deciding whether to give the money back, Sue considers the fact that the supermarket is part of a multi-billion dollar chain of stores, which is set up in such a way that no one will be appreciably worse off for her keeping the money, except perhaps the cashier, who, Sue reasons, probably deserves to be reprimanded for making such a mistake anyway. It also occurs to Sue that she is in desperate need of textbooks that she doesn't have enough money to buy, and because she is a very promising student and a good person, the world will be a better place overall if she keeps the money and uses it to buy the books. She feels that the only factor that could persuade her to give the money back would be a moral obligation to do so.¹⁵ Sue accepts the above reasoning, so she believes that if she is causally determined to keep

¹⁵ It is necessary to eliminate other potential sources for Sue's motivation to act, so that the obligation may be considered in isolation (I have more to say about this below).

the money then she has no obligation to return it.¹⁶ But, it may be argued, because Sue does not know at this point whether she is causally determined to return the money or to keep it, *and* she reasons (correctly) that if she is causally determined to return it then she can still have an obligation to do so, the obligation retains at least some action-guiding usefulness in spite of Sue's acceptance of the derivation.

My reply is as follows. It seems that there are variations of this case that must be considered in turn. One such variation relates to a range of descriptions of the case whereby Sue is more or less certain that she will be causally determined to return the money. It seems that in variations of the case in which Sue's certainty that she will return the money approaches 0, i.e., her certainty that she will be causally determined not to return the money approaches 1, the possibility (from her perspective) that she has an obligation to return it would also approach zero. If she's almost certain she will be causally determined to not return the money, then given her acceptance of the derivation from (3)-(6) above, she must be just as uncertain that she has an obligation to do so. Sue's confidence that she has this obligation would be so weak in this range of cases that the obligation could not plausibly be of any significant guidance to her.

Now take variations toward the other extreme, where her certainty that she will be causally determined to return the money approaches 1. In this range of cases Sue would be reasoning from the strength of her conviction that she will return the money to her obligation to do so. So her reasoning with respect to the action-guiding function

¹⁶ I am assuming that Sue accepts the truth of determinism. Of course, if she were convinced of the falsity of determinism the argument thus far would give her no reason to question whether she has the obligation. That would take a different argument, which I will not attempt to formulate here. For an argument that indeterminism may be incompatible with moral rightness, wrongness, and obligation, see Haji (1999, 190-202). Haji (2002, ch. 6) argues that there is one kind of libertarian theory (Alfred Mele's) that can accommodate some (although too little) deontic morality. I will not here consider the range of cases where Sue's certainty as to the truth of determinism lies at different points on a scale between 0 and 1. In order for it to be productive to do so I would first need to establish a position concerning the compatibility of moral obligation with the falsity of determinism, as well as the practical implications of such a position – issues which are beyond the scope of this chapter.

of obligations would seem to be the wrong way around, rendering the obligation useless in terms of its action-guiding function. That is, she must be fairly certain that she will return the money in order to believe that she has the obligation to do so. But in this situation it would not be the obligation that would be motivating her (directly) to return the money, but, strangely, her belief about what she will in fact do in the future based on her assessment of her current psychological condition (the strength of her belief that she will in fact return the money). If she has fully internalized the derivation from (3)-(6)¹⁷ she must be reasoning about her situation in something like the following way: “I’m only going to return this money if I have a moral obligation to do so. I accept that if determinism is true and I’m causally determined to keep the money then I have no obligation to return it. But, I feel quite strongly that I do have an obligation to return it. But why do I feel like I have this obligation? I’m an intelligent person and so I shouldn’t just assume such things. Perhaps I have the obligation because not to return the money would be wrong. But it can’t be wrong unless I have an obligation to return it,¹⁸ and I’m only obligated to return it if I will in fact return it (and I have no other reason to return it besides being obligated to do so). So does it seem likely that I will return it?” etc. In terms of guiding her action, the obligation in this situation is psychologically secondary to, because derived from, her belief about

¹⁷ Of course, implications of determinism are not the sort of thing that generally occurs to people when they are doing their grocery shopping, or at any other time. It may be psychologically near impossible, or at least very difficult, to reason in this way in practical situations. And even if it is possible it certainly seems to be very unnatural to do so. Nevertheless, we must assume that Sue is fully rational, that she does accept the derivation, and that she is prepared to subject her moral beliefs to rational scrutiny in this way.

¹⁸ I am assuming that Sue also accepts OW, which implies that her not returning the money can only be morally wrong if she is under an obligation to return it. I discuss this principle in section 1.6. Even though I have stipulated that there could be nothing other than the perceived obligation that could motivate Sue to return the money, I am happy for it to also occur to her that keeping the money would be wrong, in order to have the example be as realistic as possible (I really just want to exclude such things as fear, empathy, and conditioned responses from her reasoning). It might now be objected that it is Sue’s sense that keeping the money would be wrong, rather than her sense that she is obligated to return it, that can motivate her. I address this issue in section 1.6. Here I assume that Sue accepts what I have to say there.

what she will in fact do. So it's not really or not fundamentally the obligation that's (directly) guiding her action – it's her belief about what she's causally determined to do. The obligation here, I propose, is practically redundant, or near redundant, as it is not serving its essential action-guiding function. Sue can only believe she has the obligation to the extent that she believes she will return the money. Therefore, from a practical perspective, determinism also undermines these kinds of obligations.

This conclusion may require some further explanation. Implicit in my argument here is the claim that obligations function to motivate us to do (or refrain from doing) things that we otherwise would not be inclined to do (or to refrain from doing). I think this claim is true. This is why I stipulated in this case that there could be nothing other than the obligation, as well as perhaps the sense that keeping the money might be wrong, that could motivate Sue to return the money.¹⁹ Otherwise, consideration of her fear of getting caught, her empathy towards the cashier, her moral conditioning, etc., could cloud our judgment as to the action-guiding force of the obligation. The obligation must be considered in complete isolation from these other potentially motivating factors. I think that as a general point, when we do consider obligations in isolation from these other sorts of factors, it becomes clear that their only function in a moral system is to guide our actions. That is, they are needed in a moral system due to the fact that we don't always want to do what we believe morality requires of us. Kant (1785/1981, Ak 414) essentially makes this point when he contrasts the will of rational (human) beings with the holy (perfectly good) will: 'Therefore, no imperatives hold for the divine will, and in general for a holy will; the *ought* is here out of place, because the *would* is already of itself necessarily in

¹⁹ This is also why I claimed that if Sue's belief that she had the obligation were fairly weak (say, less than 0.5) then this belief would not be strong enough to outweigh her reasons for acting in such a way as to violate it.

agreement with the [moral] law. Consequently, imperatives are only formulas for expressing the relation of objective laws of willing in general to the subjective imperfection of the will of this or that rational being, e.g., the human will.’

When Sue is considering reasons for and against returning the money, it must be the case, in order for the example to be of any relevance, that the *only* reason she has for returning it is a perceived obligation to do so – after all, she needs the textbooks, etc. and so would prefer not to have the obligation. So if it were not that she felt the obligation fairly strongly she would have *no reason* to return the money. Considered as such in isolation, and given the fact that Sue has accepted the truth of the derivation, I think it becomes clear that in order for the obligation to serve its purpose and motivate Sue to do the thing that she otherwise would not do, she has to have a fairly high level of certainty that she does in fact have the obligation. If she lacks this high level of certainty then the obligation could not motivate her to act in accordance with it.

However, if she *does* have a level of certainty sufficient for the obligation to motivate her to act in accordance with it, then, due to the fact that there are no other factors that are motivating her to return the money (as stipulated), as well as her acceptance of the derivation, she must already be fairly certain that she will return the money. But why would she feel certain that she will return the money in this case? Given that she has good reasons not to return the money (it would enable her to buy her books, etc.) and only one (possible) reason to return it (the obligation) it should seem much more likely to her, as a prediction, that she will keep the money rather than return it. Therefore, it should seem much more likely to her that she doesn’t have the obligation. So, even if Sue starts out believing that she probably does have the obligation, and we accept that she could thus be motivated to return the money on the

strength of this belief, her consideration of the derivation should result in her questioning this belief to the point where the perceived obligation cannot do the action-guiding work that is its only purpose. It seems then, that no matter what the strength of Sue's initial conviction as to whether she has an obligation the obligation itself cannot play much of a role in guiding her actions.

A somewhat different objection might be thought to pose a problem for my claim that determinism is inconsistent with obligations that we are causally determined to satisfy. Consider, for instance, a variation of the case of Jim given above. Imagine that Jim is not merely sick, but suffering from a terminal illness, and unaware of the severity of his condition. Jim has no family, and I have been able to use this fact to persuade his doctor to share the information about his condition. The doctor is the only other person aware of Jim's impending death, but happens to be out of town for a few days on emergency medical business. I am considering whether I have an obligation to tell Jim the bad news or to keep quiet and allow his doctor to tell him in a few days (Jim is expected to live at least that long). On the one hand, I believe that as I am the only person in a position to give Jim the news right now, and as I do know him fairly well, I may be obligated to tell him myself; let's say I am about 50% certain that I have this obligation. On the other hand, I feel that under the circumstances it is only proper for Jim's doctor to tell Jim the news: I reason that the doctor was responsible for the prognosis, and information of this sort ought to be confidential between doctor and patient. Assume that I feel roughly 50% sure that I have an obligation to allow the doctor to tell Jim the news instead of doing it myself. As I only have two options in this case: to tell Jim myself or to keep quiet and allow the doctor to tell him, it might seem that either of these perceived obligations would be certain enough to guide my

actions.²⁰ For instance, if I focus on the possibility that I have an obligation to tell Jim myself, then although I am only 50% certain that I have this obligation, this degree of certainty would be sufficient to motivate me to tell him. Furthermore, it would seem that even if I am uncertain that determinism is false *and* even if I accept the proposition that if I am causally determined to violate a perceived obligation then I don't have that obligation, these considerations would be insufficient to completely undermine the action-guiding usefulness of either of these perceived obligations, i.e., either one could still motivate me to act.

I think this is a persuasive case, but I am uncertain that it constitutes a clear counterexample to my claim. With respect to my predicament in the example, if I have accepted the argument up to this point, i.e., if I have accepted that I cannot have obligations I am causally determined to violate, then I am in a position whereby it is questionable for me whether *any* perceived obligation is really an obligation. That is, anytime I find myself wondering whether I might have an obligation to perform or refrain from performing some action, in order to be consistent with my acceptance of the derivation I must view my perceived obligation as conditional on the obtaining of the future state of affairs whereby I perform (or refrain from performing) the action required to satisfy it.

So I cannot start out, in this case or any other, by asking 'Where lies my obligation?' I must begin by considering whether I have an obligation at all. Therefore, it seems that I must reason through my situation in two stages. First, I must consider the possibility that I am causally determined not to tell Jim of his

²⁰ This case is similar to the case of a moral dilemma, where it is thought that an agent can have equally compelling moral obligations to perform two mutually incompatible actions (that is, two actions that she cannot both perform). I have addressed moral dilemmas above. This case seems to be different, however, as here I am simply uncertain whether my obligation is to tell Jim or to refrain from telling him (I don't believe that I have obligations to do both).

predicament. In this case I must believe that I couldn't have an obligation to tell him, as I couldn't satisfy that obligation. Under this supposition I have very little motivation to tell him *on the basis of an obligation to do so*. Next, I must consider the possibility that I am causally determined to tell him. In this case I must believe that I couldn't have an obligation to refrain from telling him, as I couldn't satisfy that obligation. Under this supposition I have very little motivation to refrain from telling him *on the basis of an obligation to refrain*. So in each case my perceived obligation is insufficiently strong to motivate me to act. Of course, because of the nature of the case I must either tell him or not (there are no other possibilities). But, it does not follow that I must do either of these things on the basis of an obligation to do so. And I have argued that a 50% certainty that I have an obligation is insufficient to motivate me in either case (whether I am causally determined to tell him or to not tell him). I may decide to tell him because I feel an overwhelming sense of sympathy for him or because I'm tired of deliberating about the matter or because I want an answer to the question of whether or not I am causally determined to tell him. Or, if I am particularly callous, I may simply walk away and forget the whole thing. But either way, it cannot be because of any obligation that I do what I do.

It seems then, that if determinism and OIC are true, an agent cannot be under a moral obligation with respect to an action irrespective of whether or not s/he is causally determined to act in accordance with the obligation.

1.6 Moral Wrongness and Moral Obligation

In this final section I provide support for the final conclusion of the derivation (7). My purpose here is to establish that if determinism is incompatible with moral obligation then it is also incompatible with moral wrongness (and the moral rightness

of obligatory right actions) due to the interdependence relations between these notions. As noted above, non-obligatory right actions provide an exception to this interdependence, in that these are actions that are considered morally right absent any obligation to perform them. I will not here provide an argument for the incompatibility of determinism with non-obligatory right actions. This section will differ in an important respect from the preceding ones. In the preceding sections I was arguing for a more controversial claim, that the truth of determinism leads, via OIC, to the undermining of moral obligation. Here, I will be defending a more intuitively plausible claim, that an act [omission] can be morally wrong only if the agent performing it [not performing it] was under a moral obligation to not perform it [perform it]. The direction of entailment here should be noted. The other direction of entailment, that we cannot be under a moral obligation to perform [not perform] some action unless it is the case that not performing [performing] the action would be wrong, is *so* intuitively plausible, that it will not concern me. In fact it is probably analytic. I am trying to establish the general claim instanced by premise (2), which I will refer to as OW (following Haji 1999, 183):

OW: It is morally wrong for agent *S* to perform [not perform] action *A* *iff* *S* is under a moral obligation not to perform [to perform] *A*,

and the less intuitive direction of entailment (left to right) will be the focus. I will not attempt to provide a positive argument for this claim, as I think it is intuitively plausible. I will only defend the claim against a putative objection.

The objection I will consider involves an alleged counterexample. In responding to Haji, who stipulates that OW is a 'standard principle of moral obligation,' Derk Pereboom (2001, 147) describes a case in which he thinks it is

plausible that it is morally wrong for an agent to perform some action, yet it is not the case that the agent has a moral obligation to refrain:

‘For example, suppose you say to an animal-abuser, “You ought not to abuse that animal,” but then you find out that he has a psychological condition (which he could have done nothing to prevent) that makes animal-abusing irresistible for him, so that he cannot help but abuse the animal. From my point of view, there is an appreciably strong pull to admitting that the “ought” judgment was false, but there is relatively little to denying that abusing the animal is morally wrong for him.’

I agree with Pereboom that it seems more reasonable to claim that abusing the animal is wrong for the animal abuser, than it does to claim that despite his condition he is under a moral obligation to refrain from abusing. Nevertheless, I take this as merely a *prima facie* judgment, and I think there are considerations compatible with the truth of OW that may explain why we mistakenly feel that OW fails here.

First, our judgments about compulsive behavior are not straightforward. It is not clear that we believe that abusing animals can really be irresistible for this man, or similarly, that drinking really is irresistible for the alcoholic, or stealing for the kleptomaniac, etc. (I have more to say below about why this is important.) Second, there are important differences in connotations between the moral concepts of wrongness and obligation. Although this is a subtle point, it seems that judgments of moral wrongness are of a more general character, or seem to apply in a more general or universal way, than judgments of moral obligation. Very often we are thought to have obligations due to some agreement or contract we have made, or responsibility we have willingly taken on, whereas judgments of moral wrongness have more of a categorical feel: they seem to apply to types of actions, such as murder or animal

abuse, irrespective of any particular circumstances in which agents perform these actions. Of course, we do recognize certain ‘standing’ moral obligations such as obligations to refrain from lying, stealing, and so on – obligations that seem to apply categorically. On the flip-side, it is also true that there are certain actions we consider morally wrong only within some specific circumstances and not generally. Actions that serve to break promises are good examples of this kind of morally wrong action. So are actions that fail to account for specific needs or aversions of people with whom we interact. For example, showing a picture of a spider to someone or offering meat to someone would not be considered wrong unless we know (or have not bothered to find out) that the person in question is an arachnophobe or a moral vegetarian, respectively. Nevertheless, it does seem that the connotations of these notions differ: in contrast to wrong actions there are many more obligatory actions that are only obligatory because of the specific circumstances in which they are performed. So when we are presented with a case wherein the notions of wrongness and obligation both apply, but where any mention of specific arrangements (contract, promise, etc.) are absent, such as in Pereboom’s example, we feel less inclined to maintain the judgment of obligation when an element like compulsion is introduced, than we do the judgment of wrongness.

So because we feel that instances of animal abuse are always simply (categorically) wrong, because we see these as falling under an action-type that we strongly feel to be categorically wrong, we are less likely to think that such judgments are undermined by (possibly dubious) claims about irresistible psychological compulsion. An upshot of this is that in cases of actions where we would ordinarily or prereflectively feel that the moral notions of wrongness and obligation both apply, it takes a much stronger and more convincing reason to persuade us to relinquish our

judgment that the action is wrong than it does to persuade us to relinquish our judgment that the agent performing the action is under an obligation not to perform it. Whether someone is under an obligation to act or refrain from acting in some way is an issue about which we feel less certain or less strongly, as obligation seems to be a less categorical notion than wrongness, i.e., our judgments about it seem more subject to fluctuation according to contingencies presented in particular cases. Thus we are more likely to have our intuitions compromised by certain features of cases that exploit this uncertainty, just like Pereboom's use of psychological compulsion.

I think that the abovementioned cases of the arachnophobe and the vegetarian serve to give us some insight into why we might think there is a problem with OW. It seems that in these cases the actions in question are (prereflectively) certainly wrong, but it's not really clear in what sense the agents performing them were under obligations to refrain. Nevertheless, there is a third consideration that seems relevant. I think we also need to distinguish between two distinct kinds of *implications* of the notion of obligation. The case of promise-breaking can be usefully contrasted with the cases of the arachnophobe and the vegetarian (and the animal abuser) in order to highlight this distinction. It may even be useful, due to the difference in implications, to label these different kinds of obligations 'type-1' and 'type-2,' although the distinction here is subtle and intuitive – referring to (perhaps unrecognized) differences in implications and not to differences between types of action *per se*. I will refer to type-1 obligations as those for which the judgment of obligation is logically prior to any other form of moral assessment of actions. Promise-breaking seems a paradigm example of this kind of obligation. When a person performs some action that amounts to the breaking of a promise, all other forms of moral assessment of the person and the action are plausibly derived from, or dependent upon, the judgment that

an obligation has been violated. So the action of breaking the promise is *wrong* and the promise breaker *bad*, but *only because* the obligation has been violated. The fact that an obligation has been violated, it seems, is the only basis for these other moral judgments.²¹ On the other hand, type-2 obligations have quite different implications. These obligations are plausibly dependent upon other forms of moral assessment, and it is in cases of these kinds of obligations that OW seems less certain. With respect to the vegetarian example, if I know that a person is disgusted and offended at the idea of eating meat, then the intuitive force of the judgment that it is wrong of me to offer her meat is strong and obvious. But in what sense, if at all, am I under an obligation to refrain? It seems that my obligation in this case is only derivative, i.e., I am under an obligation to refrain from offering meat to a vegetarian *only because* doing so would be wrong, and I have a standing obligation to refrain from performing actions that are wrong. Yet am I under this obligation, even though it is merely derivative? Of course I am (barring, of course, considerations of determinism and OIC). At least it initially seems much more plausible that I am than that I am not, and I think that a good reason would need to be given for the rejection of the claim that I am.

I think the reason we may be uncertain about the claim that I am under an obligation in the vegetarian case is that our judgment in this respect is made from

²¹ In making these claims and distinctions I am assuming the perspective of commonsense morality rather than a revisionist view such as utilitarianism. For a relatively unsophisticated act-utilitarian, for example, all and only obligatory actions are the particular ones, given the circumstances, that maximize utility, so keeping a promise (or refraining from breaking one) will never be obligatory *per se*. Nevertheless, my distinction between type-1 and type-2 obligations could still be recognized by such a utilitarian, even though it would not relate directly to action-types such as promise-breaking, and may in fact not be a very illuminating distinction to recognize. That is, for an act-utilitarian who performs an action that fails to maximize utility, it's unclear whether that action is wrong in virtue of the fact that the utilitarian has a standing obligation to maximize utility (thus violating her type-1 obligation), or whether it's wrong, due to the fact that it fails to maximize utility, and the utilitarian has violated her (in this case, type-2) obligation to maximize utility in virtue of the fact that she has performed a wrong action. Either way, assuming the animal abuser is failing to maximize utility I don't think such a utilitarian view can help to shed much light on whether he is under an obligation to refrain (although I think it may be difficult for a utilitarian to deny that he is under an obligation to refrain, assuming that the notion of obligation plays a role in her moral outlook at all).

under the shadow of the implications of type-1 obligations, i.e., obligations that are the primary, non-derivative form of moral assessment in the cases to which they apply (as in the case of promise-breaking). When we have this kind of obligation in mind, perhaps unconsciously, we are tempted to think that the other, weaker, derivative kind of obligation is, perhaps by comparison, not really an obligation at all. But, I suggest, we should appreciate that type-2 obligations are no less obligations because they are derivative and dependent on other forms of moral assessment, and thus intuitively less strong or obvious than type-1 obligations.

Bearing this distinction between type-1 and type-2 obligations in mind, what do we now think of the case of the animal abuser? Perhaps if we temporarily remove the element of compulsion from the example we may get a clearer picture. He is not under a type-1 obligation, as neither the wrongness of his action nor our negative assessment of his character seem to be derived from an obligation on his part to refrain from abusing; we may assume that he has made no explicit promise or contract to refrain from abusing (he is not, e.g., a veterinarian). But is he nonetheless under a type-2 (derivative) obligation? If it were not for the fact that he is psychologically compelled to abuse (or that determinism and OIC are true) there would be no reason to think that he was not under this kind of obligation, as, I propose, we are always under obligations to refrain from performing actions that would be wrong to perform. So why should the fact of his compulsion leave unaffected the judgment that his action is wrong, while calling into question the judgment that he was under a (type-2) obligation to refrain? As in the vegetarian case, we may have the impression that the obligation, being a type-2 obligation, is somewhat weak, due to our (perhaps unconsciously) contrasting it with the more robust and obvious type-1 obligations and the strong intuition that, at least in general, abusing animals is morally wrong. I think

this is right. This kind of obligation *is* weaker, as it is derived from the wrongness of the action via the principle that we are always under obligations to refrain from performing actions that would be wrong to perform (as long as we are *able* to refrain, of course). Nevertheless, if I am right about type-2 obligations, in that they apply derivatively to any and all actions that are wrong, then if the wrongness of the action remains despite the compulsion, the type-2 obligation does as well, and the only way this kind of obligation can be cancelled is if the action is not wrong.

Finally, if it is reasonable to conclude on the basis of the arguments presented thus far, that determinism undermines the wrongness of all particular actions of a certain type, then it is reasonable to conclude that it also undermines the wrongness of the relevant type. For example, if it is the case that no particular action that involves the deliberate killing of an innocent human being can be morally wrong if determinism is true, then I take it to follow that it is not the case that murder is wrong if determinism is true. My analysis of Pereboom's example turns on the intuitive difference between moral wrongness and moral obligation, my claim being that we tend to think of the concept of wrong actions as applying more generally or universally than the concept of actions we are obligated to perform/not perform. This is perhaps another reason why we might be skeptical of OW (which states that an action is wrong only if we are under an obligation to refrain from performing it) when we apply it to the example. But I think this discrepancy between OW and our judgment about the example can be ironed out if we attend to the relevant intuitions and beliefs involved. Even though we generally tend to think of wrong actions in terms of abstract types of actions that are wrong – e.g., murder, abusing animals – upon reflection, if we were to think about why we consider these types wrong, we should come to realize that we can only justify thinking the type wrong on the basis of

our belief that all the instances of the type are wrong. In this way, even though we may prereflectively consider the abstract type of action ‘abusing animals’ to be wrong *simpliciter* and irrespective of obligation, upon reflection I think we should come to see that the only reason this type could be wrong is that all the instances of it are, and if I am right about type-2 obligations, then in all the instances it seems that the agent in question is under an obligation to refrain.

In this section I presented several considerations that come to bear on why we may be inclined to think that OW fails in Pereboom’s animal abuser example. One concerns the differences in *connotations* between the concepts of wrongness and obligation, my claim being that wrongness is more of a categorical notion, having connotations of generality and universality, whereas obligation has connotations of specificity and contingency. Another concerns the differences in two kinds of *implication* of the notion of obligation. Type-1 obligations imply the logical priority of the obligation over any other moral assessment of the actions to which they apply, i.e., any moral judgment other than that of obligation is derived from this judgment. By contrast, type-2 obligations imply a dependence of the judgment of obligation on some other moral judgment – in this case, a judgment of wrongness. Finally, and building on these two considerations, it seems that even though we generally think of wrong actions (as opposed to obligatory actions) as abstract *types* of actions, I think that with a little reflection we should come to see that we can only justify thinking that the types are wrong by means of our belief that all the instances of them are wrong, and all the instances, it seems, carry with them type-2 obligations to refrain. So if any action is wrong, there will be a corresponding obligation (albeit a relatively weak one) on the part of the agent to refrain from performing it.

I have now argued that an action cannot be morally wrong unless the agent performing it is under an obligation to refrain from performing it. From this claim it follows that the truth of determinism, combined with OIC, undermines judgments of moral wrongness of actions.

1.7 Conclusion

The arguments presented in this chapter have been aimed at illustrating some of the implications of determinism for what is perhaps the most fundamental and important of the moral notions: wrongness. My overall aim has been to try to establish that determinism undermines moral wrongness. The argument given on page 2 showed that if 'ought' implies 'can,' we can never genuinely be under moral obligations that we are causally determined to violate. After defending OIC from objections, I argued that due to the action-guiding function of moral obligations and the fact that we need to be confident that we have them in order for them to play this role, we have good reason to think that we can also not be under moral obligations that we are causally determined to satisfy. In the final section I defended the final conclusion of the initial argument concerning moral wrongness. I argued that judgments of all actions, right and wrong, that we believe ourselves to be under obligations to perform or refrain from performing, are straightforwardly undermined if the initial argument is correct. These include all wrong actions and obligatory right actions.

REFERENCES

- Chisolm, Roderick M. (1982). 'Human Freedom and the Self', in Gary Watson (ed.), *Free Will*, New York: Oxford University Press.
- Copp, David (2003). "'Ought' Implies 'Can'", Blameworthiness, and the Principle of Alternate Possibilities', in David Widerker and Michael McKenna (eds), *Moral Responsibility and Alternative Possibilities*, Burlington: Ashgate.
- Fischer, John (2003). "'Ought-Implies-Can', Causal Determinism and Moral Responsibility', *Analysis* v. 63, pp. 247-8.
- Frankfurt, Harry (1969). 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* 66.
- Graham, Pete, "'Ought' Does Not Imply 'Can'" (unpublished).
- Haji, Ishtiyaque (2002). *Deontic Morality and Control*, New York: Cambridge University Press.
- Haji, Ishtiyaque (1999). 'Moral Anchors and Control', *Canadian Journal of Philosophy* v. 29 no. 2.
- Honderich, Ted (1988). *A Theory of Determinism*, New York: Oxford University Press.
- Kant, Immanuel (1785/1981). *Grounding for the Metaphysics of Morals*, trans. James W. Ellington, Indianapolis: Hackett Publishing Company.
- Lewis, David (1986). 'The Paradoxes of Time Travel', in David Lewis, *Philosophical Papers II*, New York: Oxford University Press.
- Pereboom, Derk (2001). *Living Without Free Will*, New York: Cambridge University Press.

Strawson, Peter (2003). 'Freedom and Resentment', in Gary Watson (ed.), *Free Will* (2nd ed.), New York: Oxford University Press.

van Inwagen, Peter (1983). *An Essay on Free Will*, New York: Oxford University Press.

CHAPTER 2

HOW BELIEF IN DETERMINISM THREATENS RATIONAL DELIBERATION

2.1 Introduction

Can an agent deliberate rationally while believing that her process of deliberation is causally determined by factors beyond her control? There are reasons to think that this should not be possible. While helping us to better understand why belief in determinism might conflict with deliberation, I will argue that these reasons also constitute compelling evidence for the conclusion that it in fact does so. I begin by considering several recent accounts according to which rational deliberation is compatible with the agent's belief that she is causally determined. I contend that these accounts, while largely successful given one specific characterization of the problem, do not address the core issues from which it arises. Drawing on arguments from Thomas Nagel and Brian O'Shaughnessy, I respond to this compatibilist position by developing two complementary proposals for identifying these underlying core issues, which concern the motivation, sense of control, and psychological integration required for rational deliberation. I argue that an appreciation of these resulting requirements for rational deliberation makes it plausible that they cannot be met by an agent who believes that her deliberative process is causally determined.

2.2 Motivating the Problem

Aristotle was perhaps the first to identify the challenge to the compatibility of determinism with deliberation. He claims that if 'everything is and happens of necessity...there would be no need to deliberate or to take trouble (thinking that if we do this, this will happen, but if we do not, it will not)' (Aristotle, *De Interpretatione*

18b26). Relatedly, the ‘lazy’ or ‘idle’ argument, which appears to have originated at least as early as the Stoic Chrysippus in the third century BCE, runs as follows. ‘If it is fated for you to recover from this disease, then you will recover, whether you call in a doctor or not; similarly, if it is fated for you not to recover from this disease, then you will not recover, whether you call in a doctor or not. But one or the other is fated; so there is no point in calling in a doctor.’ Generalized, and interpreting ‘fate’ as fate-determinism¹ the argument appears to show that if determinism is true there is no point in doing anything. According to Cicero, ‘if we obeyed this we would do nothing at all in life’ (Cicero, *On Fate* 28-9). Although the conclusion of the argument concerns motivation for action rather than deliberation, it would seem that if there’s no point in doing anything then there’s no point in *deliberating* about doing anything. I argue below that deliberation is an active process that requires effort and motivation, so if the lazy argument applies to action in general, it would apply to deliberation as well. More recently, Richard Taylor (1966, 182) has argued that ‘if a man believes, concerning some of the actions he is going to perform, that there already exist conditions causally sufficient for his performing them, and conditions which therefore render them inevitable, then he cannot deliberate whether or not to perform them.’²

¹ For Chrysippus, ‘fate’ seems to incorporate both the modern mechanistic aspect of causal determinism as well as a teleological element involving the ‘one Cause’ or the ‘Reason of the universe.’ The relations between the notions of fate, providence, necessity, and causal determinism in Stoic philosophy are quite complex, but the one thread that seems to tie most of these elements together (with the possible exception of the teleological ones) is inevitability, which would seem to be implied both by ‘determinism’ as well as the central notions of ‘fate’ (cf. Bobzien 1998, ch. 1; Long and Sedley 1987, 340, 342-343). Although the inclusion in the lazy argument of the clause ‘whether you call the doctor or not’ suggests localized rather than global inevitability (i.e., it is only your living or dying that is inevitable rather than whether or not you call the doctor, or anything else) this does not matter a great deal for the purposes of this chapter. I include this argument as background primarily to indicate that the general concern I am discussing here is not new.

² It will be apparent that while both the lazy argument and the quotation from Aristotle seem to concern inevitability per se, my focus, as indicated by Taylor’s claim, is on the implications for a deliberator of believing that such inevitability applies to her deliberative situation. It may be noted, however, that both of the earlier quotations could readily be interpreted in terms of the epistemic issue, as Cicero seems to have done with respect to the lazy argument.

Thomas Nagel provides a particularly illuminating characterization of the problem, although in a broader framework. He draws our attention to two seemingly irreconcilable perspectives we can adopt toward our actions. On the one hand, we naturally (and, he thinks, unavoidably) assume an *internal* standpoint when we act. Our sense of our own autonomy takes the form of certain beliefs we hold about ourselves and our relation to the world. We believe, for instance, that we are not causally determined to act in the ways we do by antecedent circumstances, whether these involve facts about the world or about us. We believe that we generally have a range of open possibilities available to us when we act, and when, by acting, we make one of those possibilities actual, the final explanation of why we did what we did is to be found only in our intentions, thereby our justifying reasons and purposes (Nagel 1986, 115). We are, we might say, the ultimate source of our actions.

In contrast to this internal standpoint we are also able to adopt an *external* view of our actions whereby we see them objectively, from the point of view of an observer. From this perspective we see ‘not only the circumstances of action as they present themselves to the agent, but also the conditions and influences lying behind the action, including the complete nature of the agent himself’ (113). Viewed this way, an action seems to happen *to us* rather than being something *we do*. Whether or not we think of our actions as causally determined, this view allows no room for any explanation of action that is not causal. So, viewing ourselves and our actions merely as causes and effects, ‘we cease to face the world and instead become parts of it; we and our lives are seen as products and manifestations of the world as a whole’ (114). Importantly, Nagel thinks that this objective view ‘produces a sense of impotence and futility with respect to what we do ourselves’ (112). Although he doesn’t specifically discuss the implications for deliberation, it would not involve an injustice to him to extend his

point here in the following way. When we adopt this view, our attitude toward the objects of our deliberation tends to take on an aspect of ambivalence or resignation, perhaps appropriately described as the generalized conclusion of the lazy argument, that is, “What’s the point in doing anything?” ‘The sense that we are being carried along by the universe like small pieces of flotsam’ (112) describes the *spectatorial* rather than participant attitude we take toward our own lives that results from taking an external point of view, and this seems to come into conflict with some of the prereflective assumptions we hold while engaging in a process of deliberation.

2.3 Belief in the Availability of Multiple Possibilities for Action

One feature noticeably absent from this picture is an explanation of *why* this drastic change in our attitudes occurs, or in Nagel’s words, ‘what intelligible belief [about our autonomy] is undermined by the external view’ (117). What is it about the internal perspective that is lost when we view our actions naturalistically – as simply parts of causal processes? What exactly is it about deliberation in particular that seems imperiled by the belief that the deliberative process and its results are determined? The one strand that seems common among the descriptions of the problem given above is that when we view an action from an external standpoint, or more specifically, as causally determined,³ we can no longer regard it as the realization of one out of a range of actions genuinely available to us. Peter van Inwagen (1983, 155) characterizes this aspect of deliberation in the following way. ‘[I]f someone deliberates about whether to do A or to do B, it follows that his behavior manifests a belief that it is *possible* for him to do A – that he *can* do A, that he has it within his power to do A – and a belief that it is possible for him to do B.’ But, it is argued, if

³ According to Nagel, thinking of our actions as determined is not the only way to produce the effect of viewing them from the external perspective, but it is the easiest (110).

determinism is true these beliefs are false: it is never within our power to perform an action other than the one we are causally determined to perform. So it might seem that an essential feature of deliberation is lost if determinism is true. As Taylor explains,

...the thesis of determinism entails that, in the case of any action that any man ever performs, there are conditions antecedent to his action which render it causally impossible for him to perform any other. Deliberation on the other hand, presupposes that an action which one contemplates doing, and concerning which he deliberates, has been rendered neither causally impossible nor causally inevitable by any conditions obtaining at the time of deliberation (Taylor 1966, 182).

The claim is that engaging in a process of deliberation requires a belief that one genuinely has alternative possibilities for action. Belief that causal determinism is true is apt to conflict with this, since one will believe of any of the alternatives one is considering while deliberating, that it is either eliminated as a genuine possibility, or will occur inevitably, excluding any other possibility. Thus one will believe that only one alternative one considers while deliberating is genuinely open, which would contradict a belief required for deliberation given Taylor's view.

2.4 Deliberation-Compatibilism

In order to resist this conclusion, deliberation-compatibilists such as Tomis Kapitan (1986), Dana Nelkin (2004), and Derk Pereboom (2008)⁴ advocate replacing the requirement of a belief in genuine or metaphysical alternatives on the part of the deliberating agent with an epistemic openness and consistency requirement which, it is

⁴ See also Philip Pettit (1989) and John Searle (2001).

claimed, is possible for agents to satisfy even if they believe determinism to be true.

Pereboom's formulation is:

(S) In order to deliberate rationally among distinct actions $A_1 \dots A_n$, for each A_i , S cannot be certain of the proposition that she will do A_i , nor of the proposition that she will not do A_i ; and either (a) the proposition that she will do A_i is consistent with every proposition that, in the present context, is settled for her, or (b) if it is inconsistent with some such proposition, she cannot believe that it is,

and,

a proposition is 'settled' for an agent just in case she believes it and disregards any doubt she has that it is true, e.g., for the purpose of deliberation (Pereboom 2008, 294).

These authors agree, however, that epistemic openness by itself is not sufficient for the possibility of deliberation, as van Inwagen's 'two-door' example shows. Here an agent finds himself in a room with only two doors, one of which he believes to be unlocked, the other locked and impassable, although he does not know which is which. Van Inwagen thinks the reason this man is unable to deliberate about which door to leave by is that I can't deliberate about whether to perform a certain action unless I believe it is possible for me to perform it (van Inwagen 1983, 154). Here it seems as though the fact that the agent believes his options to be open in the sense required by (S) – he has no certain beliefs about what he will do, and each possibility he imagines is consistent with what is settled for him – is insufficient to allow him to deliberate about which door to leave by.⁵ Pereboom's solution, following suggestions by Kapitan and Nelkin,

⁵ Both Nelkin and Pereboom agree that, although the agent can't deliberate about which door to open, he can deliberate about which door to *try* to open (Pereboom 2008, 297). I do not think this is clearly the case, however, as it seems to me as though genuine deliberation requires that the deliberator have, or at least believe himself to be capable of finding, some evaluative basis for differentiating between his options. Van Inwagen's example is not described in much detail, but if we imagine that there is no

is to supplement the openness condition with a condition that captures the agent's belief that his deliberation will be efficacious in bringing about each action he is deliberating about whether to perform – in this case, the opening of each door. To this end Pereboom offers:

(DE) In order to rationally deliberate about whether to do A1 or A2, where A1 and A2 are distinct actions, an agent must believe that if as a result of her deliberating about whether to do A1 or A2 she were to judge that it would be best to do A1, then, under normal conditions, she would also, on the basis of this deliberation, do A1; and similarly for A2 (Pereboom 2008, 299).

As Pereboom points out, this condition is not met in the two-door example, but it is met in any normal case of deliberation even where the deliberating agent believes the truth of determinism and thus believes that she lacks genuine alternative possibilities.

2.5 Interpreting the Problem

In the discussion so far I have been assuming that the question of whether belief in determinism is compatible with deliberation is plausibly interpreted as a question of whether it is possible for a rational agent to deliberate and to believe that she lacks multiple alternative possibilities for action. It seems that Pereboom's analysis, involving (S) and (DE), adequately answers this question.

However, perhaps there are beliefs essential to the process of deliberation other than those concerning alternative possibilities, and these beliefs are inconsistent with belief in determinism. Then the deliberation-compatibilist conditions for rational

detectable difference between the two doors such that the agent could find a non-arbitrary reason, based on some evaluative distinction, for *trying* one door instead of the other, the mental process he would be engaged in would not properly be called deliberation. (See Pereboom (2008, 298) for an opposing view.) Some of what I have to say in section 2.6 should clarify this point.

deliberation, although adequate for answering the formulation of the problem in terms of alternative possibilities, might nonetheless be missing something that is not captured by this formulation. I will now argue that there are such additional beliefs, and they concern one's motivation and sense of control. The lazy argument provides a clue: what *is* the point of trying to work out what to do when you believe that whatever you will do is already fixed and unalterable due to causally determining factors beyond your control, and you are therefore not its ultimate source?⁶ But before attempting a more precise formulation, we need to be clear about what properly counts as deliberation and what does not.

2.6 What is Deliberation?

Let us begin with a rough characterization of the essence of deliberation: it involves a rational agent's working out what to do from a range of perceived alternatives. This already captures several of the important characteristics. First, deliberation, often referred to as 'practical reasoning,' differs from theoretical reasoning in that its goal, or at least its end result, is intention-formation and action rather than simply belief-formation. When reasoning theoretically I am trying to work out what to believe; when reasoning practically I might be trying to work out what to believe, but I am also, and essentially, trying to work out what to do. Second, deliberation is a rational process undertaken by a rational agent. Perhaps this characteristic is not obvious, although for practical reasons I think it's best to limit discussion of deliberation in the present context to rational deliberation. Finally, the

⁶ Given my disagreement with both the deliberation-compatibilists *and* deliberation-incompatibilists such as Taylor and van Inwagen, that the basis of the problem is the incompatibility of belief in determinism with a deliberating agent's sense of being the ultimate source of her actions, rather than her believing she has alternative possibilities for action, it would not be inaccurate, following current usage, to refer to my position as 'source deliberation-incompatibilism' in contrast to what we might call the 'leeway deliberation-incompatibilism' of Taylor and van Inwagen.

essential subject-matter of deliberation is always perceived alternatives for action and their consequences, whether these alternatives involve simply performing or not performing a particular action, or performing one or a subset out of a range of actions not all of which can be performed. Now, as the deliberation-compatibilist analysis showed, a deliberating agent need not believe that each of her perceived alternatives is a genuine possibility for her, but she must at least be considering each of these alternatives as options for acting – otherwise, she would have nothing to deliberate about.

This initial characterization clearly does not capture all that's essential to deliberation, but before trying to refine the account further it will be useful to rule out of consideration candidates for deliberation that should not properly be considered instances of it. From the above points alone we can already see that deliberation is distinct from some other psychological processes in that it is *active* (Pereboom 2008, 291; Taylor 1966, 168-9). In contrast to, say, daydreaming, reminiscing, worrying, remembering (as opposed to trying to remember), or enjoying a good sunset, deliberating is something that requires conscious effort and sustained concentration. It is also *purposeful* in that it has a decision to act as its goal (Taylor 1966, 168), and it clearly involves a strong sense of involvement in, and *control* over, both the process and the outcome of the process (the resulting action) on the part of the deliberating agent. Accordingly, we can exclude as candidates for deliberation certain psychological processes such as prediction, vicarious 'deliberation,' vacillation, and retrospective and hypothetical 'deliberation.'

(a) Mere *prediction* is excluded on the grounds that a sense of control over the predicted outcome is absent.⁷ It seems that the extent to which our attitude toward an outcome is merely predictive we have no sense of involvement in the process leading to the outcome. We are simply observers, perhaps involved to the extent that we might be adversely or favorably affected by the outcome, but lacking any ability to control or affect the outcome itself. (b) *Vicarious deliberation*, or deliberating about what someone else should do, can also not properly be considered deliberation. We may, for example, ‘deliberate’ about what a character in a movie or novel should do, realizing all the while that not only are we incapable of affecting what happens, but that the events in the story have already been written and played out – we are merely spectators with incomplete knowledge of how the events will unfold.⁸

Simply vacillating between competing inclinations is excluded due to the vacillating agent’s not deliberately adducing and weighing reasons for perceived alternative actions, and to his decision’s being based only on the presence of his strongest desire at the time he makes it (Taylor 1966, 169-70). Here we may have a sense of involvement and even of control over our decision, but the process does not engage our reason: we are not weighing reasons for various actions against each other; we are simply acting in an unmediated way on present desires. We can assume that

⁷ David Velleman (1989, 143) argues that the basis of deliberation is in part theoretical, and in part involves motives by which we theorize about our own actions. These motives induce us to seek hypotheses that are true of our actions, and also to perform actions that are true to our hypotheses. The motives incline us to do whatever we have predicted. So we get ourselves to do things by predicting that we will do them. These resulting predictions, if self-describing, constitute intentions to act. Velleman’s view is plausible, but I do not believe it threatens my account, as the predictions he has in mind as part of deliberative processes are not the relatively simple predictions I want to exclude from the category of deliberative processes, even if such predictions take the agent’s own actions as objects.

⁸ In some cases of vicarious deliberation we may feel as though we have some control over how another agent will act (say we are trying to help a friend decide what to do), but we recognize that the actions the other agent performs are always ultimately under the control of that agent rather than us, and our sense of involvement is present only to the extent that we think we can convince, coerce, or otherwise motivate the other agent to do what we want him to do. In this way vicarious deliberation amounts to little more than a combination of projection and prediction.

many non-human animals vacillate in something like this way, but are not capable of deliberating. Galen Strawson provides a useful illustration of this process of vacillation using the example of a dog choosing between jumping into one of two channels of a river in order to save either his master or his mistress (Strawson 1986, 141-2). Retrospective deliberation, or deliberation about what one should have done when the time for action has passed, is not possible, again primarily due to the fact that a sense of control over the outcome cannot be present, as well as a realization that one is powerless to affect or alter what has already happened.

Finally, hypothetical deliberation might take the form of either conditional deliberation or deliberation ‘off-line.’ Conditional deliberation occurs when a person considers what to do *if* a certain situation obtained in which a decision would be required. I think this is difficult to distinguish from genuine deliberation in any principled way, and in some cases may properly be thought of as such. This may be distinguished from deliberation off-line, where a person may deliberate *as if* in a situation where deliberation is required. This is not genuine deliberation in cases where the options considered are too far removed from the deliberator’s actual situation for him to feel a sense of personal involvement in the process and to form an intention relevant to those options as a result of the process. But again, some cases of off-line deliberation may be hard to distinguish from genuine deliberation, especially where the situation a person is deliberating about is considered likely to arise. In such cases it may be that a person forms a conditional intention that can be called upon should the type of situation actually arise. Perhaps vicarious deliberation could in some ways be considered a subcategory of off-line deliberation.

Collecting the results of the previous three sections we can characterize deliberation as follows:

Deliberation is a practical, rational process engaged in by a rational agent in response to the recognition of future practical uncertainty relevant to that agent's actual situation, that causes or activates a felt need to act so as to ensure that the uncertainty is resolved in a way favorable to the agent.⁹ Its subject-matter is perceived alternatives for action and their projected outcomes, and the process, which essentially involves adducing, comparing, and evaluating reasons for and against these perceived alternative actions and outcomes, carries with it a sense of the agent's personal involvement in the process and outcome, a sense that the process and outcome are under the agent's control, and a belief that the deliberative process will be causally efficacious in bringing about the agent's preferred alternative, whatever that should turn out to be.

2.7 Motivation and Control

I have already stated my endorsement of the deliberation-compatibilists' resolution of the issue of the compatibility of belief in determinism and its evident consequences with the central epistemic conditions required for deliberation. Since I agree with the deliberation-compatibilists that a determinist deliberator need not believe that it's possible for her to perform each of the alternative actions she's considering during a process of deliberation, I have conceded that such a deliberator need not have inconsistent beliefs when deliberating.¹⁰ We may now ask whether the

⁹ I do not intend this to exclude altruistic motives. An agent might find it most favorable, for example, that agents other than her benefit from whatever she decides to do, perhaps even at great personal cost to her.

¹⁰ I concede this point reluctantly. From a practical, psychological standpoint it seems intuitively implausible that a person could deliberate over a range of actions whilst lacking a belief that it's possible for her to perform each. However, I am unable to substantiate this claim directly. The remainder of this chapter can in some respects be seen as an attempt to put pressure on the deliberation-compatibilists' claim that a requirement of belief in genuine alternative possibilities is unnecessary for deliberation.

refined characterization of deliberation provides a clue as to what it is about deliberation that is threatened when the deliberative process and its results are believed by the deliberator to be causally determined by factors beyond the agent's control.¹¹

Earlier I mentioned that what's threatened most likely concerns motivation and a sense of control on the part of the deliberating agent. So perhaps the refined formulation might help to explain this. The aspect of deliberation, formulated as above, that seems most relevant to the agent's motivational state involves the agent's feeling a need to act so as to resolve her perceived practical uncertainty in a way favorable to her. So, could a belief that the process and outcome of deliberating are determined interfere with this motivational aspect of the process in a way that compromises the agent's ability to deliberate?

One possibility that stands out is that when we are conceiving of our deliberation and its result as determined we take on, to some extent, the attitude of a predictor or perhaps a vicarious deliberator. This change in attitude involves a sense of a split between what we might think of as our Cartesian selves and the other psychological elements involved in deliberation.¹² One aspect of vicarious

¹¹ It may be thought that I have overlooked a simple solution to this compatibility problem, which is that the reason a person has difficulty deliberating while simultaneously considering her deliberation as causally determined is that deliberation requires too much concentration to be undertaken while at the same time applying an abstract notion like inevitability to one's situation. I think this is mistaken, however. First, the inevitability implied by determinism is not particularly difficult to grasp, and second, it is fairly easy to deliberate while bearing some other relatively simple metaphysical proposition in mind such as that God does or does not exist or that one might be a brain in a vat. I do believe it is very difficult to conceive of one's ongoing thoughts as causally determined, at least for very long, but depending on how deep such meditation goes I don't see that this is necessarily an issue having to do with limitations on one's ability to concentrate. Some of the explanations I give below of the difficulties of applying a belief in determinism to one's deliberation seem to also apply to active non-deliberative thought, but I don't go into any detail about this issue as it is not my primary focus.

¹² In using the term 'Cartesian self' I am departing somewhat from Descartes' usage in his Second Meditation. Here he attributes to the self all that cannot be doubted as illusory, for example, affirmation and denial, willing, imagination, and sensory perceptions (Descartes 1641/1996, 28). In using the term I am intending to denote a more minimal conception of the self that excludes any mental function,

deliberation, as described above, seems relevant: when we begin to view ourselves and our deliberative activity as merely parts of the deterministic causal nexus – as inevitable and unalterable – it may be that the sense of control we normally take ourselves to exert over our weighing of reasons and intention-formation is compromised, perhaps seen as illusory.^{13 14} We would take on what I described above, interpreting some of Nagel’s remarks, as a spectatorial attitude, much like the attitude we naturally adopt toward a character in a movie. We would predict what a character might do given what we take to be his beliefs and desires, and perhaps try to deliberate for him, or in his place, to weigh up the pros and cons of the various actions we imagine are available to him and work out the best thing for him to do, or the thing we would consider best to do were we in his situation. All the while we understand that we are unable to influence any of his future actions or any of the events that are inevitably going to unfold on the screen.

So even though motivational aspects remain to some extent in a case of vicarious deliberation, our activity is more akin to theoretical reasoning than practical reasoning, which can be seen in the fact that the process by itself cannot cause the

characteristic, property, etc., that is inessential to a basic awareness of one’s existence, or to whatever it is that is required for there to be, in Nagel’s terms, something it is like to be a thing (cf. Nagel 1974). As will be seen to be of particular relevance below, willing is excluded.

¹³ Below I claim that intentions are never willed (s. 2.8, n. 17). I do, however, believe that we ordinarily assume that we will our intentions, which is why considering them as determined can compromise the sense of control we feel we have over them, and thereby (to some extent) the whole deliberative process.

¹⁴ I am not ignoring the possibility that a determinist deliberator may endorse a kind of control that is compatible with determinism, such as John Fischer and Mark Ravizza’s ‘guidance control’ (see Fischer and Ravizza 1998, ch. 2). But the point here concerns the *sense* of control one ordinarily takes oneself to have when one deliberates, and even though I am unable to directly argue for it here, it is not plausible to me that one’s sense of one’s control when one deliberates allows being able to do only what one is causally determined to do. Fischer and Ravizza in fact claim that having what they call ‘regulative’ control (which involves alternative possibilities) at certain important points in our lives is the ‘intuitive and natural picture of ourselves’ (28). Whether it is psychologically possible to deliberate on the basis of a belief in guidance control alone, and not regulative control, is a further question. Considerations raised in the remainder of this chapter should make it plausible that this would be insufficient.

formation of an intention to act. This could be due to differences in motivation involved in each process. In a case of vicarious deliberation we can be motivated to engage in a reasoning process that superficially resembles genuine deliberation in the sense that we recognize a practical uncertainty and are motivated to consider various alternative courses of action, adduce reasons for each, weigh these against one another, and come to a conclusion about what would be best to do. But all of these activities are of a quite distinct nature from their genuinely deliberative counterparts. First, my recognition of uncertainty cannot really be said to be the recognition of practical uncertainty on my own part: I am not motivated to work out what *I* should do but what *someone else* should do. This would be a key difference between the two processes. The end result of a reasoning process (be it intention-formation or merely belief-formation) would be determined by the kind of motivation that sets the process in motion. If I initially engaged in a process motivated by the thought “What should Sally do?” or “What would I do if I were Sally?” the resulting process by itself could only end with a resolution of this kind of question, i.e., “That’s what Sally should do” or “That’s what I would do if I were Sally.” These kinds of thoughts clearly do not express intentions (we cannot intend vicariously). The reason vicarious deliberation does not produce intentions, again, concerns control. I realize that I do not control the actions of the character in the movie, and I also realize, on a very basic level, that the scope of things a person can intend is limited to exactly the scope of things that (she believes) she can do. Brian O’Shaughnessy makes a similar point, albeit in a different context. In formulating an account of the conditions that must be met in order for an act to be intentional, he writes, ‘an intentional act must be the exercise of a power that is significantly contributory to the success of trying,’ and that ‘to perform an intentional [A]-doing...one must at least harbour an intention of *trying to do* [A]’ (O’Shaughnessy 1980, 326). An act is not intentional unless it is an exercise of a

power one takes oneself to have, and that one is exercising in an attempt to succeed at doing what one intends to do (325-27).^{15 16}

2.8 Causal Efficacy

There will be an important difference between deliberation under the assumption of determinism on the one hand, and these cases of prediction and vicarious deliberation on the other, and it concerns causal efficacy. Staying with the experience of watching a movie as an illustration of both prediction and vicarious deliberation, we can see that there is something missing here that is present in a case

¹⁵ I am assuming that believing that one has the power to do *X* is roughly equivalent to believing that doing *X* is under one's control, and that, at least for a rational agent, trying to do *X* entails believing that one has the power to do *X*, or at least has a high probability of succeeding at doing *X* if one tries.

¹⁶ This section and the previous one involve claims that highlight certain fundamental differences between the processes of theoretical and practical reasoning, perhaps the most important being that intentions can't be formed solely from theoretical reasoning, and that the motivation needed for practical reasoning requires a sense of control that is incompatible with believing the evident consequences of determinism. These claims are not uncontroversial. For example, Brian O'Shaughnessy's account of deciding subsumes the processes of theoretical and practical reasoning under the one category of activity. He thinks that the event that completes a process of theoretical reasoning, which he refers to as 'deciding whether,' and the event that completes a process of practical reasoning ('deciding to do'), although necessarily distinct from one another, are 'reached along the same road' (O'Shaughnessy 1980, 300-301). He gives only one example to substantiate this claim, which I believe can be plausibly interpreted in a way that preserves the distinction in question, although I am unable to go into any detail here. Furthermore, my claim, that purely theoretical reasoning can't cause intentions, is lent some plausibility from David Hume's views about the causes of actions, wherein he argues that 'reason alone can never be a motive to any action of the will... and that it can never oppose passion in the direction of the will' (Hume 1739/2000, 2.3.3.1). Indeed, my claim could be seen as a corollary of this so long as 1) practical reasoning always proceeds on the basis of some 'passion' (desires count as passions on Hume's account); and 2) what I have been referring to as 'theoretical reasoning' is no more than a function of the understanding, in Hume's sense. 1 seems to be true if the above formulation of deliberation is accurate, at least in the sense that what motivates a process of deliberation is a desire ('felt need') to find a favorable resolution to a practical uncertainty. 2 also seems true if we consider Hume's explanation of the role of the understanding. He thinks it works in two different ways: it informs us about 'the abstract relations of our ideas' as in areas such as mathematics and mechanics (2.3.3.2); and it informs us about 'those relations of objects, of which experience only gives us information,' in other words, the causal relations between physical objects and processes (2.3.3.3). The second of these perhaps best describes the kind of reasoning I have been referring to as 'vicarious deliberation' in that it involves the ascertaining of causal factors (e.g., the subject's beliefs and desires) for the purpose of prediction, in order to derive a judgment about what would be best for the subject to do.

of genuine deliberation, even when the deliberator is considering her deliberative experience as determined. When we genuinely deliberate we believe, correctly, something like Pereboom's (DE), that is, we believe that we, by way of our deliberation, are directly involved in the causal production of our action – we will cause it by way of our deliberation, whatever it is. In addition to the abovementioned fact about intention-formation, as well as differences in an agent's sense of control over the process, this belief in efficacy is what most prominently distinguishes genuine deliberation from prediction and vicarious deliberation. When we 'deliberate' vicariously we understand that nothing we might think or do can have any effect on the outcome – we can only, for instance, hope for the best and wait to see what happens. So it might be objected that even if believing one's deliberation to be determined robs one of a sense of control, this does not reduce one to the role of spectator, since one can still believe, correctly, that one's deliberation will be causally efficacious in producing the action. So this absence of a sense of control would not result in the drastic change in attitude I am trying to explain.

To answer this concern, we need to examine the notion of causal efficacy more closely. If we strip the term 'efficacy' of any of the connotations that don't belong to it when it is qualified by 'causal,' we are left with the bare notion of something that causes. These connotations include, for instance, 'effective' and 'efficient,' and may involve notions such as ability, capability, accomplishment, purpose, competence, usefulness, desire-fulfillment, and importantly, control. In the present context then, for one's deliberation to be causally efficacious is just for it to cause something – most immediately and obviously, an intention to act. So is the belief that our deliberation will likely cause an intention to act significant enough to prevent the thought that our deliberation is causally determined from reducing our deliberative attitude to that of a

spectator? It certainly adds something that's missing in prediction and vicarious deliberation.

But given the would-be determinist deliberator's belief that the deliberative process she is experiencing is inevitable and thus ultimately beyond her control, as is whatever intention is formed as a result, it seems to me that a belief that her deliberation will cause an intention would neither prevent the onset of, nor diminish the effects of, a spectatorial attitude. An analogy may serve to illustrate this point. Imagine a normal process of deliberation where options are adduced and weighed, and an intention is formed as a result. Now consider Stacy, who is about to engage in such a process but becomes aware that whenever she is about to begin to weigh reasons for and against each option, a chip that has been implanted in her brain activates and alters the strengths of her desires relevant to the weighing process, although after she begins to weigh reasons the chip becomes inactive. So, for instance, if the decision to be made concerns a choice between ordering Chinese or Italian food for dinner, she knows that if she begins to try to find reasons for and against each alternative, the chip will activate and, say, strengthen her fondness of tomatoes and her dislike of reading through long menus. Stacy would believe correctly that her reasoning is efficacious here, but I propose that her sense that her deliberation is under her control would be compromised to the extent that she would find it very difficult to form an intention about what to do with respect to this particular decision.¹⁷ Although rather fanciful,

¹⁷ This seems to imply that intentions are willed. I side with Brian O'Shaughnessy in thinking that neither beliefs nor intentions are directly willed: 'Thus, while a man may do what he hopes may bring about his deciding to do [roughly, forming an intention], deciding to do is not and never can be an activity. In this sense we may say that, while the will can strew incentives before itself, it cannot directly determine its own direction. Indeed, were it to be able to do so, the flood-gates would break and the will carry all before it: mind, reason, sanity, world itself!' (O'Shaughnessy 1980, 298, 301). Some libertarians (agent-causalists and non-causalists in particular) are likely to resist the claim that intentions are not directly willed. Although I think the claim is plausible, I am not able to defend it from such accounts here. I refer briefly to these accounts in the final section of my third chapter, and I discuss intention-formation in section 3.7. I think the claim applies in an interesting way to this case. It seems plausible that the reason Stacy would have trouble forming an intention about which kind of food to

this example indicates that a belief in the causal efficacy of one's deliberation would not by itself help, or not help very much, to prevent or reduce the effects of believing one's deliberation to be (determined and) ultimately beyond one's control.

There is something more specific to say about causal efficacy. It seems that the above list of some of the potential connotations of 'efficacy' might usefully be expanded by the addition of one more notion that does not belong to it in the present context – that of *making a difference*. It is natural to be motivated to deliberate on the basis of a supposition that if you don't deliberate, or don't deliberate correctly or well enough, then you will bring about a less than favorable outcome. We might formulate a supposition of this kind as follows:

DM: I am able, via a process of deliberation, to bring about an outcome that is more favorable to me than the outcomes that would result if I either do not deliberate or do not deliberate well.

Some thought or belief like DM arguably underlies the motivational aspect of deliberation mentioned earlier, i.e., the felt need to act so as to resolve a practical uncertainty in a favorable way. It seems plausible that a process of deliberation would

order is that she would realize that any such intention would in some sense not really be her own, due to her belief that the desires that helped form it had been tampered with. So given that she is hungry it would be natural for her to try to form an intention directly, somehow bypassing the effects of the chip. If O'Shaughnessy is right she would be unable to do this. So the only options available to her would seem to be to choose at 'random,' to vacillate, or to submit to her desire for Italian food. Were she to do any of these, however, the intention that she forms would not be the result of a process of rational deliberation. Therefore, I think it's plausible that Stacy's belief that her deliberation with respect to this decision (should she attempt to engage in it) will be causally efficacious would not be sufficient to prevent her from adopting a spectatorial viewpoint toward it, and thus not sufficient for her to be able to engage in it. And I think the best explanation for this is that her sense of control over the entire process has been compromised. Of course, it is not compromised in the same way that I'm claiming one's sense of control would be compromised when one is considering a process of deliberation to be causally determined. For instance, Stacy would realize that the outcome of her decision may not entirely or accurately reflect her propositional attitudes (she might not 'really' like tomatoes as much as she now seems to). However, the point of the example is to highlight the limitations of the role that belief in the causal efficacy of one's deliberation can play in allowing one to engage in deliberation when one's sense of control over the process has been compromised. So as long as the agent's sense of control *has* been compromised, and to a similar extent in both cases, it's immaterial how this has happened.

never begin (and if already underway, would cease) unless the agent were motivated to deliberate by a belief specified by a conditional of this form. And if the truth of DM is incompatible with a belief in determinism, then perhaps this is something that a determinist deliberator recognizes on some level, and another part of the explanation of why deliberation could be prevented or disrupted when one believes it to be determined.¹⁸ It would provide a further reason to think that a belief in the causal efficacy of her deliberation would not help to extricate a would-be determinist deliberator from her predicament, and allow her to be motivated in the right way to deliberate. That is, the belief that one's deliberation is causally efficacious would not compensate, in a motivationally-relevant sense, for the sense of loss of control resulting from thinking of one's deliberation as determined, if what is lying behind the notion of causal efficacy is this difference-making assumption.

So, is a belief of the form specified by DM compatible with the agent's applying her belief in determinism to her deliberative process? Under scrutiny, it might well not be. Take an example. Let's say I've been watching an advertisement about vacations in Florida and it occurs to me that I have some time off in a few weeks and it might be fun to go. But then it occurs to me that I really should stay in town because my sister is due to have a baby right around that time and would probably be grateful for my moral support. This is, among other things, my recognition of practical

¹⁸ It might be objected that I am ignoring what would be a compatibilist reading of 'making a difference,' whereby an agent's belief in determinism would not be incompatible with her belief that, in deliberating, she can make a difference. Certainly, there is a sense of 'making a difference' that is quite compatible with the truth of determinism. We might, for instance, correctly say that my installing deadlocks on the doors made the difference between the burglar's gaining entry or not, in that if I had not installed the locks, the burglar would have gained entry. However, I believe that these kinds of conditionals, although compatible with the truth of determinism, are subject to the same kind of problems that conditional analyses of 'can' and ability are subject to. For a refutation of these kind of conditional analyses, see Chisholm (1964). Furthermore, I believe that a suitably reflective determinist will appreciate that the compatibilist conditional reading of 'making a difference' is inapplicable to her situation, when she is considering how she is capable of affecting future states of affairs given her belief in determinism.

uncertainty that motivates me to begin to deliberate. The next step in a normal deliberative process would seem to involve my imagining various aspects of my perceived options: in this case, to go to Florida or stay home. Under normal circumstances I would weigh reasons for and against each option, perhaps arrive at some new beliefs along the way, and eventually form an intention either to go or to stay. My motivation for weighing reasons for and against each option is to at least a significant extent, if not entirely, based on an underlying or standing desire to act so as to bring about the most favorable option available to myself, my sister, and anyone else who might be affected, and to avoid any less favorable option. This type of thought or belief, again, is grounded in DM: “I’d better think hard and make sure I go over all the possibilities, because *if* I miss something or misjudge something, then things will end up worse than they otherwise might.”¹⁹

Now let’s run through the process again, this time assuming that I believe determinism to be true and that I’m applying this belief to my deliberation. First, as I’m watching the advertisement about Florida and thinking that it looks like a fun place to be, I remember that I have some time off in a few weeks and also that my sister is due to give birth around that time. So an uncertainty relevant to my future predicament is recognized. At this point it occurs to me that determinism is true,²⁰ so it

¹⁹ I have been referring to these goings on in the minds of deliberators as thoughts *or* beliefs due to my not wanting to commit myself to their being consciously recognized. I certainly don’t think that the sentence in quotation marks here is something that persons ordinarily consciously think when they are embarking on a process of deliberation. But I do think that it, or something like it, is something deliberators believe, perhaps merely dispositionally or sub-doxastically, even if it is never consciously thought.

²⁰ To a certain extent my stipulation that the truth of determinism occurs to the agent at this particular point in the deliberative process is arbitrary. It seems that the thought could occur to the agent at some other point and the effects would be similar, although if it occurred too late in the process the chances are that the agent might have already done enough weighing of perceived options to be fairly certain what to do, and so it would be harder to see exactly what the effects of the application of a belief in determinism would be. The question of whether being in a state of intending to do something is compatible with believing determinism to be true and applying this belief to one’s intention in some way is beyond what I can address here.

is either inevitable that I go to Florida or inevitable that I not go. I'm far from certain which is determined to be the case, and it's also true that both my going and my not going are consistent with everything else that's settled for me. I also believe that my deliberation and decision, whichever way it goes, will be causally efficacious in bringing about either my going or staying (I thus satisfy Pereboom's conditions (S) and (DE)). But when I begin to imagine each scenario and to consider reasons for and against each one, I think that all these thoughts I'm now having were determined to occur to me just as they are doing, as are the thoughts I'll be having two minutes from now, and two minutes from then, right up until I make my decision about whether or not to go. In fact, so I think, even before I was born it was determined that I would be trying to make this decision about whether or not to go to Florida *and* analyzing the process in exactly the way I am. So ultimately, I reason, none of this is really under my control – it just feels that way, or at least it did.

So, what happens to my motivation to weigh reasons for and against each option in this scenario? It seems that whatever motivation remains would only be of a kind that motivates theoretical reasoning. This would not get me very far along the way to forming an intention about whether or not to go to Florida, as theoretical reasoning, I have claimed, won't produce an intention. For the idea of trying to ensure that I ultimately do the thing that would be most favorable and avoid doing anything less favorable (a notion that relies on the truth of DM) cannot plausibly motivate me to work out what to do when I believe that whatever I think and whatever I ultimately do was determined to be just as it is and just as it will be. I do realize that my deliberation, or whatever other thoughts occur to me, may well be causally efficacious in relation to my ultimately going or staying. But as we have seen, there is nothing more to a thought's being causally efficacious than its causing something. And my

belief that my thoughts may play a causal role in my future actions would not by itself be likely to provide me with enough sense of control over the process to enable me to genuinely deliberate, when I believe that either going or else staying is already fixed and unavoidable. The richly motivational thought described by DM would make little sense to me in this context, and so is unlikely to be available as a motivator. It doesn't seem plausible that the thought, "I'd better think about this carefully or things will end up worse than they otherwise might," can motivate me to deliberate when I believe that whatever I'll do, as well as all the thoughts and other events that cause me to end up doing it, are already fixed. The motivating thought relies on the assumption that I, as deliberating self, am the ultimate source of control over which option will be realized. But if I believe that whatever will come to pass with respect to the subject of my uncertainty is inevitable, then my belief in such difference-making control would seem to be undermined: there is really no difference to be made. So the potentially motivating thought that relies on the assumption that I have this kind of control should also be unavailable to me.

2.9 Integration and Detachment

I have now provided one account of why applying a belief in determinism to one's deliberation seems problematic. It explains why two factors that are psychologically required for rational deliberation – a sense of a particular kind of control and the right kind of motivation – would appear to be incompatible with a belief that one's deliberation and its results are determined and inevitable. However, I don't think we yet have the full picture. Perhaps a process of deliberation could be undertaken and completed on the basis of some other kind of motivation, and without the sense of difference-making control I have claimed is required. Perhaps we could

be suitably motivated to deliberate on the basis of a sense of curiosity as to what we are causally determined to do, even while recognizing that we lack the control required to ensure a favorable outcome and avoid a less favorable one. In an attempt to rule out such a possibility I will provide a complementary account, which is intended both to illustrate further interesting features of the psychological bases of deliberation, and to show the extent to which believing one's deliberation to be determined can interfere with the process. The basis of the account is the observation that a certain level of psychological integration is required for deliberation. I will explain below how the two accounts are related.

What is psychological integration? Nagel gives us at least a starting point. According to him, when we begin to adopt the external view of ourselves and our actions, we subject more and more aspects of ourselves to scrutiny – our desires, beliefs, feelings, impulses, motives, principles, habits, and the biological, psychological, and social factors that form them (Nagel 1986, 119, 114). Of particular relevance in the present context is his claim that as more and more of the self is 'swallowed up in the circumstances of action,' the less the self is able to act. Taken to the extreme, he thinks this process of self-examination has dire consequences:

At the end of the path that seems to lead to freedom and knowledge lie skepticism and helplessness. We can act only from inside the world, but when we see ourselves from outside, the autonomy we experience from inside appears as an illusion, and we who are looking from outside cannot act at all (119-20).

In order to understand this point, and how it relates to determinism and deliberation, we need to look in more detail at this distinction between being 'inside' and 'outside.' From the internal perspective I experience my feelings, desires,

impulses, habits, etc., as well as my body (generally), as bound together in a single, integrated entity whose attention, when I think and act, is directed outward, toward things ‘external’ to me (be those physical objects, sensations, ideas, etc.) that are not experienced at that time as part of me as thinker, and that I can apprehend, adopt attitudes toward, alter, and influence – the objects of my thoughts and the purposes and goals of my actions. By contrast, one of the distinguishing features of the external perspective is that when I begin to subject myself to the scrutiny to which Nagel refers, the complex, multifaceted, unified entity that I prereflectively take to be my self ‘shrinks’ to exclude every mental item from which it can be distinguished and separated. It is in this way that I come to distance myself from all these mental items and consider them in the way I ordinarily consider things outside my skin.²¹ It does not seem that we need to be focused constantly on these mental items in order to occupy this perspective; what is essential to the external perspective is only that state of mind whereby the ‘other’ side of the self-other distinction includes as many aspects of our selves as possible, and the ‘self’ side includes as few as possible – this minimal ‘self’ side is what I above referred to as the Cartesian self.

Although it may not be necessary to be focused on aspects of one’s self in order to occupy this perspective, perhaps a useful way of understanding the distinction between the internal and external perspectives is to consider what happens when we try to observe ourselves acting. In cases of very simple and familiar actions, such as wiggling my toes, I am able to perform this action and observe my performance of it at the same time. I am able to observe the act of wiggling being performed, just as I

²¹ It will be apparent that the metaphor I am employing here does not align precisely with Nagel’s. Where he illustrates this distancing effect by describing the self’s viewing these mental items from a position external to them, I am describing it in terms of the self’s shrinking away from these items. Although both metaphors capture the same basic idea (the sense of distancing), the ‘shrinking’ description is somewhat more accurate for my purposes, as it is more congenial to the idea that the distancing occurs in degrees.

would observe someone else's performing similar movements, while at the same time feeling as though it is me who is causing and controlling it; I am, in a sense, inside and outside of the action simultaneously. Nevertheless, genuinely observing oneself acting seems impossible. As O'Shaughnessy explains:

If one is to relate as observer to anything then one has to be 'without' it, whereas if one is intentionally to do anything then one has to be 'within' it. Now either we remain 'within' the action we are attempting to observe, in which case we may have a completely empty and self-delusive experience of observation – comparable to Wittgenstein's example of the right hand attempting to pay the left hand money, or else we remain 'without' in some more or less serious sense and genuinely seem to observe the action. But, remaining 'without', we lose the action as ours in gaining the observation: we lose any 'withinness'. The action becomes for us a mere event in the world, and we ourselves become dispersed and lost amongst the bric-a-brac of the world: we become of the world in our own eyes: we suffer the experience of loss of identity. (I say 'experience' because nobody can actually *lose* his identity) (O'Shaughnessy 1980, 32).²²

²² If this is correct, then why do we sometimes feel as if we are observing ourselves acting? I think the explanation is that the impression we can have of being an observer of our actions can only occur in cases where the action in question is one (or one of a kind) that we have performed so many times in the past that we are able, once we have initiated the action, to continue or complete it automatically. But what seems fairly clear is that at any particular time during the performance of an action, we can occupy only one of these two perspectives with respect to it. We can be consciously willing ourselves to act, and at that time occupying the internal perspective with respect to the action, i.e., we are more or less fully integrated as actor and our focus is directed away from ourselves, perhaps toward the goal of the action. Or else the action is being performed automatically, and we are able to occupy the external perspective with respect to its performance. In the latter case we are 'detached' from the action to the extent that we are able to perceive its occurrence from the point of view of observer. In a sense, our attention is still being directed outward in such a case, although it is directed toward something that is ordinarily (in typical cases of acting) experienced as an integrated part of our self – hence my description of the self as shrinking.

With some grasp of the distinction between the internal and external perspectives, we are in a position to see if this distinction is of any help in understanding why considering one's deliberation to be causally determined is problematic. Two issues require our attention. One is how, and to what extent, conceiving of our thoughts and actions as determined places us in an external position with respect to them. The other is whether deliberation is possible from this position. As we have seen, according to Nagel, thinking of our thoughts and actions as determined does indeed seem to place us on the outside, in the position of spectator with respect to them. Exactly why this happens is so far unclear, as is the extent to which this reaction is universal. I suspect that when it does occur, it is the result of a sense that something fundamental to agency has been lost, or is at least temporarily unavailable or inaccessible. Above I argued that considering the implications of determinism might well rob us of the usual sense of control we feel ourselves to have over our thoughts and actions, thereby rendering a kind of motivation that is essential for deliberation unavailable (that of trying to ensure a more favorable outcome and avoid a less favorable one). But there is something else about the internal perspective that may be lost, and it primarily concerns the subject-object and the self-other distinctions just canvassed.

When thoughts occur to us as psychologically integrated agents, we do not consciously focus on the fact that they do. Rather than thinking about, for example, what kinds of thoughts we are having, why they are occurring to us, what thoughts are coming next, whether we want to be having them, etc., we focus exclusively on their contents. As I've described it, our focus is directed outwards, toward things we experience as external to ourselves as thinker and actor. By contrast, when we are conceiving of our thoughts as determined and inevitable, our gaze tends to turn

inwards, toward the thoughts themselves. This is no doubt due to our recognition that the idea that *everything* is determined and inevitable, including all our thoughts, comes into conflict with the natural, prereflective belief we seem to have, that we are, in Nagel's words, not completely determined to act in the ways we do by antecedent circumstances, and that the final explanation of why we do what we do is to be found in our intentions, and thereby in our justifying reasons and purposes. We might describe this prereflective view of ourselves as encompassing the idea that we are the *ultimate cause* or *ultimate source* of our thoughts (more about this below). Even if compatibilism about determinism and free will is correct, it seems that our initial reaction to the consideration that our thoughts, and in particular our active, purposive thought-processes, are inevitably caused to be as they are by circumstances ultimately beyond our control, can alter the normal, prereflective perspective we have toward them. It is plausible that this change in perspective is caused by a conflict between our idea of inevitability and our sense of being the ultimate cause of our thoughts. On the one hand, we have a sense of the implications of determinism and their application to our own thoughts, e.g., that each thought we have is merely an inevitable consequence of its causal antecedents. On the other, there is the sense we ordinarily have, when viewing things from the internal perspective, that when we are engaged in an active, purposive thought-process (as opposed to, say, passively observing or daydreaming) we, as psychologically integrated thinkers, create the new thoughts we have, in the sense that we are their ultimate source. Central to this view of ourselves is the impression that when we think and act we are something over and above a mere collection of thoughts – thoughts which are simply caused by other thoughts, and ultimately by things other than thoughts; we are an independent entity who *has* thoughts, and yet is fundamentally distinct from them.²³ Our thoughts seem, after all,

²³ This description seems to cohere fairly well with the view of the self described by agent-causal

to be things we can view as objects, from a perspective external to them. This kind of view of ourselves could be strongly challenged by the belief, implied by the truth of determinism, that all the thoughts that occur to us do so inevitably, and were determined to be exactly as they are, and exactly as they will be, since before we were born.

A natural reaction to this apparent incompatibility, then, is to mentally ‘withdraw’ into that sense of being or self, which we prereflectively assume to be distinct from our thoughts, which are now considered to be ultimately beyond our control and alien to us. So even if, due to our considering our thoughts as determined, the self that we ordinarily experience as constitutive of our identity has almost completely shrunk due to the self-observation brought on by such considerations, we still believe we are essentially an entity distinct from any thoughts we might have. But our experience reveals to us that we can only withdraw to a certain extent, as ultimately, there seems to be little, if anything beyond our thoughts into which to withdraw.²⁴ In this state, each time we are aware of a new thought occurring to us we immediately become alienated from it as we attempt to distance ourselves from it. Even if we only get to this rather deep point once during such an exercise of self-reflection, the experience leaves us, while we are mindful of the implications of

theorists. Despite differences between agent-causal views, their proponents all seem to agree that the self that causes thoughts and actions is an entity distinct from, and irreducible to, mental events such as the thoughts to which I have been referring. The self as agent-cause is a temporally enduring entity (as opposed to a mere collection of temporally fleeting mental states and events) which, despite its ability to directly cause free acts, is not itself caused (or, at least, not causally determined) to act as it does by prior states and events. Prominent recent examples of such accounts are Timothy O’Connor (2000), and Randolph Clarke (2003). Although my descriptions of our prereflective sense of our own agency do reflect some of the views of agent-causalists, I do not intend such descriptions to indicate my endorsement of causation of events by substances (the agent-causalist’s central claim) nor any details of particular agent-causal accounts.

²⁴ There is, one would think, an obvious limit to the extent to which we can view our own thoughts as objects. Aside from the potential regress that would be involved if we were to try to ‘step back’ from every thought we have and focus on it from an external perspective, there is also the threat of what we might call psychological disintegration, or what O’Shaughnessy refers to as the experience of loss of identity.

determinism, in the strange position of feeling as though we are detached from, and external to, our own thoughts.

Given these reasons why considering our thoughts as determined can place us in an essentially external position with respect to them, we now need to see whether deliberation is possible from this perspective. The account above, for why we can't observe ourselves acting, is also relevant here. Suppose it's true, as I've argued, that the explanation of why we can't do this stems from the requirement that, when we intentionally act, we occupy the internal perspective in relation to our action. Then in order to find out whether we can deliberate from this perspective we need only to know whether the process of deliberating shares those characteristics of intentional action that are the basis of this practical incompatibility. I gave toe-wiggling as an example of an action that, while it seemingly can be performed from the external perspective, really only seems this way due to the action's being of a kind that is so familiar and simple that we can get the impression that we are consciously willing our performance of the action at times when this is actually not the case. I think we can see more clearly why this can't be the case if we look at a more complex action. Writing is an apt example. Sometimes, as we are writing, we seem able to view our activity from an external perspective, even if only for an instant. This seems to be because, just as in the toe-wiggling case, we have performed this kind of action many times before, and so we are able, once we have initiated the action of writing a particular word or sentence, to carry the action to completion automatically. We are even able to perform some other complex intentional actions at the same time, such as talking. But there are limits to this. We cannot, for example, complete a word that is relatively unfamiliar to us, or a sentence with relatively complex grammatical structure, unless we are, in O'Shaughnessy's terms, 'within' the action. And this holds

for more than just actions. It seems that in order to engage in any intentional activity of any complexity or novelty requires that our relation to such activity be internal. Relating as observer to any such activity seems for us psychologically impossible. Examples bear this out: think of, say, composing a grocery list or trying to remember a dream. Deliberation, one of the activities that requires the most from us due to its complexity and novelty (we cannot ‘rehearse’ a process of deliberation and ‘perform’ it again later), would be no exception to this rule. Our relation to our reasons for acting needs to be an internal one. If we are detached from such reasons, viewing them from the perspective of observer, then they cannot – via the production of an intention – motivate us to act.

In summary, the Florida example in the previous section was intended to show how applying a belief in determinism to one’s deliberative situation can result in the loss of a sense of control over one’s deliberative process. This would render a motivational element essential to deliberation unavailable: the belief that one can, by way of one’s deliberation, ensure a favorable outcome and avoid an unfavorable one (DM), or more fundamentally, that one can make a difference to how things will turn out. Without this key motivational element, I claimed, one’s reasoning is restricted to the theoretical.

Then in this section I considered the notion of psychological integration, and argued that believing one’s deliberation to be determined can result in the adoption of an external view toward one’s thoughts and actions – a position from which one is unable to deliberate. These are the essential points of my proposal:

- (a) Viewing our own thoughts as determined and ultimately beyond our control conflicts with the prereflective assumption that we are the ultimate source of our thoughts, and thus distinct from them.

- (b) We react to this apparent incompatibility by attempting to withdraw from our thoughts as they are occurring to us – to retreat into the distinct entity in which we assume ourselves to essentially consist.
- (c) The position we find ourselves occupying if we attempt to distance ourselves from our own thoughts in this way is the same as the one we occupy when we attempt to observe our own actions, especially the complex ones. The common factor is that both deliberating and deliberately acting have the requirement of ‘withinness.’ But our relation to such activities must be an internal one, and so once we are distanced, we can no longer deliberately act or deliberate.

As these two accounts of the effects of viewing one’s deliberation as determined – one focusing on difference-making control, the other on detachment and the internal-external distinction – do not obviously coincide, I shall now attempt to reconcile them into a single account.²⁵

2.10 A Unified Account – Control, Motivation, Detachment, and Dissociation

The first step toward reconciling these accounts relates to the initial change in attitude toward one’s deliberation that can occur when one views it as determined, which I have described as a sense that one’s control over the process and the outcome is compromised and perhaps even seen as illusory. It seems plausible that when we

²⁵ The fact that I have given these two, largely distinct accounts may seem problematic, as it seems evident that even theoretical reasoning requires that an internal relation hold between the subject and the process. So if both practical and theoretical reasoning require this internal relation then, it may be argued, either it’s not the case that considering one’s deliberation as determined results in the adoption of a completely external perspective, or it’s not the case that we are unable to carry out complex and novel reasoning processes when trying to adopt this perspective, or I’m wrong to claim that theoretical reasoning is still available to one who is viewing their thoughts and actions as determined. However, I think some of this can be cleared up if we pay more attention to detail and make some distinctions that have previously been ignored. The unified account I present here is essentially an attempt to do this.

begin to view our deliberation as determined, we immediately feel as though we are in control of less than we had previously assumed – that is, our range of control seems diminished. How deeply this loss would be felt depends on how thoroughly we are applying these implications of determinism to our situation, and perhaps also on whether we have any preconceived compatibilist beliefs or tendencies toward such beliefs. At the very least we are likely to feel immediately that we lack control over some anticipated distinct event such as the outcome of our expected deliberation – whether we will, for example, end up deciding to go to Florida or not. I think the explanation of why it is the imagined outcome of our deliberation that is most immediately affected in this way, rather than, say, our present thoughts, is that we are more clearly able to conceive of them and distance ourselves from the former. The proposition ‘I will end up in Florida for my vacation or I will not,’ is a straightforward disjunction with simple content about future states of affairs, whereas applying the notion of inevitability to the thoughts we are now having or are about to have is more difficult. It’s a very strange exercise, and requires some concentration, to try to conceive of the very next thought that will occur to you as determined, and to keep applying this idea to new thoughts as they occur to you. But the distinct proposition that it is determined and inevitable that you will either go to Florida for your vacation or not go, is, like the proposition that it is determined whether you will recover or die from an illness (the lazy argument), relatively simple to apprehend.

Now it seems possible that an agent would cease to apply the concept of determinism to anything further at this point. Such an agent may have come to believe that his vacationing in Florida is not an open possibility – it will either inevitably occur or inevitably not occur – but he has subjected no other aspect of his life to his belief that determinism is true (perhaps he has become a kind of localized fatalist with

respect to his vacation). But even here it appears plausible that he would lack the motivation to deliberate about whether to go to Florida. For an essential motivating element of deliberation, the belief or presupposition that you must deliberate or deliberate well so as to ensure a favorable outcome and avoid an unfavorable one, seems to be psychologically incompatible with the belief that whether you will go or not is inevitable given past states of the world and the laws of nature. It is plausible that while the psyche of such an agent is basically integrated in the sense I have described, his impaired sense of control over this particular decision, and the resulting loss of the motivation required for deliberation, places him in a spectatorial role with respect to the decision as to whether to go to Florida. He can still engage in a process of prediction as to whether he will end up in Florida, and it seems that he may be able to reason about this decision to a certain extent, in the sense that he can identify reasons for and against, and even evaluate them against one another. But even if such processes do not result in any degree of psychological disintegration, that is, even if he does not start to question or otherwise examine his relevant desires, beliefs, or thought-processes, something will be missing. These thoughts are restricted to what I have been calling theoretical reasoning, and alone cannot cause him to form an intention to go to Florida or an intention not to go, while he is maintaining his belief that whether he goes is determined. He lacks the sense of control required to fully engage in these processes in the way that he must for them to allow him to form intentions. He can predict, vacillate, flip a coin, or try to persuade himself that determinism is false, but cannot deliberate about whether to go to Florida while conceiving of either his going or not going as determined and inevitable.

This is perhaps the least intrusive way in which belief that determinism is true could interfere with one's deliberation. In this case the agent is in a limited spectatorial

role, his psyche can remain basically integrated, and his ability to reason is only impaired to the extent that he is unable to reason through a particular decision as a deliberating agent. But just as spectators in general, though by definition lacking control over what they are observing while they are in a spectatorial role, still feel as though they control their own thoughts, the agent in this example feels a loss of control only with respect to something quite external to his psyche – the decision to go to Florida or not. But what might happen when an agent considers the implications of determinism for his situation more thoroughly?

If our agent is reflective enough to consider the implications of determinism, not just for his decision as to whether to go to Florida for his vacation, but for his thoughts and actions generally, he may start to feel his control compromised in a more fundamental way. The scope of what he sees to be determined is likely to broaden to include not only his future actions but his present thoughts. Due to perceiving an incompatibility between this self-conception and his prereflective view of himself as fundamentally the ultimate source of his thoughts, he may experience a kind of psychological withdrawal into a deeper self, in an attempt to distance himself from his determined and ultimately uncontrolled thoughts. From this position he would be, using Nagel's description, unable to act at all. Nor, I have claimed, is he able to deliberate. For from this (almost completely) external perspective, his own conscious thoughts would be immediately separated from himself as they occur, and the self he has withdrawn into would be unable to reason, even on a theoretical level. At the extreme, we might describe his predicament as that of being detached from his (conscious) will.

This extreme scenario is, I think, unlikely to occur for any longer than an instant due to human cognitive limitations, and the nature of the will. Velleman (1989,

173) gives a plausible characterization of the will as a locus of autonomy, which ‘contains motives by which it can restrain, redirect, and reinforce our other motives for acting, in accordance with our own conception of those motives.’ If this is accurate, then these motives of the will are not simply contingent desires, but ‘ground level’ motivational elements, that probably form the basis of all active human cognition. As such, they are likely to be highly resistant to complete detachment from the self in the way I have been describing, that is, as viewed from the external perspective, as part of the self that is uncontrolled in virtue of its nature and activity being determined and inevitable. Nevertheless, it is interesting to note that a complete split between the self and the will may be possible. Dissociative disorders are fairly common, and although unlikely to be caused by philosophical reflection alone, they provide evidence of the sort of detachment that might occur in this extreme scenario, even if only very briefly.²⁶

I have now described two extremes with respect to the predicament of a would-be determinist deliberator. The scenarios depicting these extremes represent two

²⁶ Depersonalization disorder, one of the dissociative disorders, is characterized by ‘a feeling of *detachment* or estrangement from one’s self... The individual may feel like an automaton or as if he or she is living in a dream or a movie. There may be a sensation of being an *outside observer of one’s mental processes*, one’s body, or parts of one’s body...[and] *a sensation of lacking control of one’s actions [is often present]*’ (American Psychiatric Association 1994, 488 (emphasis added)). According to *DSM-IV*, episodes of dissociation also occur with (at least) the following disorders: dissociative amnesia, dissociative fugue, dissociative identity disorder (formerly multiple personality disorder) as well as obsessive-compulsive disorder and some other anxiety disorders (417-23, 477-491). Although I am reluctant to endorse descriptions of the disorder that refer to a person’s being detached from themselves, such a description is understandable. Exactly what the self amounts to is mysterious, but given the above account of detachment, it seems plausible that if enough of what is ordinarily considered or experienced as part of the integrated self becomes detached from a deeper and more fundamental self or consciousness, a person may well have the experience of a split *within* the self, which would explain this experience of feeling that everything happening to one is really happening to somebody else and that ‘identity is disappearing.’ This detachment would also explain feelings of loss of control experienced during episodes of depersonalization and perhaps most acutely experienced by sufferers of obsessive-compulsive disorder. Although the relation holding between the self and the will is unclear, it seems possible that the conscious will, as the basic acting or striving part of the self, is inessential to it, and therefore could become fully detached also, resulting in a passive, perhaps catatonic observational state resembling the minimal conception of the Cartesian self referred to above (s. 2.7, n. 12).

magnitudes on a single continuum ranging from mild to severe impediment to deliberation. Thus the task of reconciling the two accounts previously presented (the first in sections 2.7 and 2.8; the second in section 2.9) is resolved in locating each account on this continuum. On one end of the scale the effects of applying a belief in determinism to one's deliberative situation are limited to a proposition about a particular future uncertainty: whether or not I will do *A*. In this case the subject's psyche remains integrated and 'outward-looking' but, I suggest, her lack of a sense of control over whether or not she will do *A* restricts her motivation to reason about it so as to exclude the essential difference-making element required for genuine deliberation. On the other end of the scale, a subject's reflection about her own thoughts and motivational states might well go so deep that virtually all the contents of her psyche are detached from whatever is essential to her self, possibly even including the fundamental motivational elements I have been referring to as the will.

Could it really be that applying a belief in an abstract notion like determinism to one's thoughts could result in the shutting down of one's entire conscious motivational system, leaving one essentially without a conscious will? Actually, I think this is possible. But, as I've mentioned, this is the extreme case, and the self-reflection involved at this level, due to its nature, requires a great deal of concentration, so even if a person is reflective enough to reach this point, she is not likely to remain in this will-less state for very long. Concentration is itself willed or involves willing, so there will be a point at which one's ability to further meditate on the problem will be inconsistent with the effects of such meditation. But perhaps a very temporary catatonic state of will-less awareness might ensue, immediately followed by the reactivation of the will and its reintegration into the self. Furthermore, a suitably reflective agent engaging in such a process will almost certainly fluctuate

between various levels of integration, sense of control, and motivation, and so the effects I have been describing are likely to occur only sporadically.

2.11 Conclusion

There is more than one reason to doubt that an agent can be rational and deliberate while believing this very process to be causally determined. While the deliberation-compatibilist may be correct to claim that the epistemic requirements concerning alternative possibilities can be consistently met, there are practical, psychological factors essential for deliberation that can conflict with the agent's belief that her deliberation is determined. This belief can conflict with suppositions about motivation and control required for deliberation, and with the psychological integration that deliberation demands. As noted at the outset, the accounts I have provided here are intended for two purposes. One is to help us to better understand why it is that rational deliberation has been thought to conflict with belief in determinism. But I have also hoped to make credible that once we come to appreciate the nature of this apparent conflict, it should become plausible that there is indeed a conflict, and that an agent who applies her belief in determinism to her deliberative situation carefully and thoroughly should find that she is unable to rationally deliberate.

REFERENCES

- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* 4th ed., Washington, DM, American Psychiatric Association.
- Aristotle (c. 350 BCE/1963). *De Interpretatione*, J. L. Ackrill tr., Oxford: Oxford University Press.
- Bobzien, Susanne (1998). *Determinism and Freedom in Stoic Philosophy*, New York: Oxford University Press.
- Chisholm, Roderick (1964). 'Human Freedom and the Self', in Watson (ed.), *Free Will*, New York: Oxford University Press.
- Cicero (45-44 BCE/1991). *On Fate*, R. W. Sharples tr., Warminster: Aris & Phillips Ltd.
- Clarke, Randolph (2003). *Libertarian Accounts of Free Will*, New York: Oxford University Press.
- Descartes, René (1641/1996). *Meditations on First Philosophy*, John Cottingham tr., Cambridge: Cambridge University Press.
- Fischer, John Martin, and Ravizza, Mark (1998). *Responsibility and Control*, New York: Cambridge University Press.
- Hume, David (1739/2000). *A Treatise of Human Nature*, Oxford: Oxford University Press.
- Kapitan, Tomis (1986). 'Deliberation and the Presumption of Open Alternatives,' *The Philosophical Quarterly* 36, 230-51.
- Long, A. A., and Sedley, D. N. (1987). *The Hellenistic Philosophers*, Cambridge: Cambridge University Press.

Nagel, Thomas (1974). 'What Is It Like to Be a Bat?' *The Philosophical Review* 83, 435-50.

Nagel, Thomas (1986). *The View From Nowhere*, New York: Oxford University Press.

Nelkin, Dana (2004). 'Deliberative Alternatives,' *Philosophical Topics* 32, 215-40.

O'Connor, Timothy (2000). *Persons and Causes*, New York: Oxford University Press.

O'Shaughnessy, Brian (1980). *The Will: A Dual Aspect Theory* v. 2, Cambridge: Cambridge University Press.

Pereboom, Derk (2008). 'A Compatibilist Account of the Epistemic Conditions on Rational Deliberation,' *The Journal of Ethics* 12, 287-306.

Pettit, Philip (1989). 'Determinism with Deliberation,' *Analysis* 49, 42-4.

Searle, John (2001). *Rationality in Action*, Massachusetts: MIT Press.

Strawson, Galen (1986). *Freedom and Belief*, New York: Oxford University Press.

Taylor, Richard (1966). *Action and Purpose*, Englewood Cliffs: Prentice-Hall.

van Inwagen, Peter (1983). *An Essay on Free Will*, New York: Oxford University Press.

Velleman, David (1989). *Practical Reflection*, Princeton: Princeton University Press.

CHAPTER 3
THE CONSEQUENCES OF INDETERMINISM FOR RATIONAL
DELIBERATION

3.1 Introduction

I argue in chapter 2 that applying a belief in causal determinism to our deliberative situation is likely to prevent us from being able to deliberate. I give two related reasons for this. One reason is that the perceived inevitability of what we will do robs us of a sense of control over our deliberation, resulting in the unavailability of the particular kind of motivation required for deliberation: trying to ensure a favorable outcome and avoid an unfavorable one. This puts us in the role of spectator to our own actions, a position from which only theoretical reasoning is possible. The other reason is that when we see our deliberative activity as causally determined we tend to start viewing all our thoughts as determined, and we begin to adopt a spectatorial or external viewpoint with respect to them. This is the result of a psychological conflict between the idea that our thought processes are determined and inevitable, and our prereflective belief that we are the ultimate cause of our thoughts, i.e., that we are not causally determined to think and act as we do. Once our capacity to ultimately cause our thoughts begins to be questioned, this conflict tends to force our attention to the thoughts we are having as they are occurring to us, thereby detaching them from us.¹ Due to the fact that we are unable to observe ourselves acting, when we become detached from our own thoughts in this way we should find ourselves unable to deliberate – deliberation being the kind of purposeful activity that requires us to stand in an internal psychological relation to it, i.e., we must not be detached from our

¹ ‘Detachment’ in this context describes the mental distancing that occurs – to the extent that this is possible – when our occurrent thoughts themselves become objects of our thinking.

reasons for acting. The question to be addressed here is whether a conception of free agency that posits indeterminism as an essential factor is similarly problematic with respect to our ability to deliberate. In simplest terms, can libertarian accounts of free agency be of any help in giving us back those elements essential for deliberation that a deterministic conception makes unavailable? The focus of this chapter will be restricted primarily to libertarian accounts referred to as ‘event-causal’: causal accounts of free agency that do not assume an irreducible relation between a substance (agent) and an event.²

3.2 Why a Deterministic Self-Conception is Problematic for Deliberation:

Difference-Making Control and Ultimate Control

The description of determinism given in the first premise of Peter van Inwagen’s Consequence argument,

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past (van Inwagen, 1983, 56),

is unsettling because determinism would challenge some very fundamental beliefs we hold about our capacities as free, rational agents, and our experience of intentionally thinking and acting. There are at least two such beliefs that seem to be incompatible with belief in determinism, and they both concern control. One of these, discussed in chapter 2, is our belief that in choosing and acting we are able to *make a difference* to

² My focus is restricted in this way for two reasons. One is logistical, in the sense that many of the central points in my argument, as well as its conclusion, seem to apply to all or most non-agent-causal libertarian accounts (as I explain in the concluding section), whereas agent-causal accounts would require separate treatment, which is beyond the scope of this chapter. More generally, I agree with Robert Kane’s view that libertarians ought not to appeal to ‘special kinds of entities, or special forms of agency or causation, that are invoked specifically to salvage libertarian intuitions and are not needed to account for free agency generally – whether libertarian or nonlibertarian’ (1996, 116-17). I do claim in chapter 2 (s. 2.9, n. 23) that we tend to prereflectively experience our agency in ways that somewhat reflect the agent-causalists’ central claims. But as I mention there, I do not intend to be thereby endorsing any such claims.

how things will turn out. This belief underlies a motivational element essential for deliberation, which I described as a felt need to act so as to resolve a practical uncertainty favorably. This belief in difference-making control has the consequence that there is a range of possible ways the future could be. But it implies more than this. Despite arguments that aim to show that the availability of alternatives for action is compatible with determinism (for example, Hume 1748, Ayer 1954, Lewis 1981), and despite deliberation-compatibilists' arguments that we don't need to believe that we have alternatives in order to deliberate (for example, Kapitan 1986, Nelkin 2004, Pereboom, 2008), it is plausible that when we consider all future events concerning ourselves as determined and inevitable, we will come to believe that there is no way to avoid them – that we lack the kind of control that would enable us to make a difference to how things will turn out.

The other aspect of our ordinary self-conception that seems to be incompatible with belief in determinism, which is even more clearly identified by the first premise of the consequence argument, is the belief that we have ultimate control over the initiation of our thoughts and actions, in the sense that they are not merely the extensions of causal chains, but are created solely by us, the agent, as they occur. This belief is captured well by Chisholm (1982, 32): 'Each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing – or no one – causes us to cause these events to happen.' We might describe this as the belief that we are the *ultimate source* of our thoughts and actions. However, in using this description we must be careful to distinguish the notion of the source of our actions being within us (say, in our brain) from the notion of ourselves, the agent, *being* their source. It is only the latter notion that captures the experience we have (or perhaps it's safer to say our belief about the experience) of deliberately initiating

causal chains – for example, a new line of thought, a new idea, or an intention. So, being the ultimate source of our actions requires us to have ultimate control over them. Control requires a controller. Therefore, the notion of being an ultimate source involves more than the idea that a causal chain is initiated within us; it must be initiated *by* the agent as controller, in such a way that it is not merely a causal consequence of previous thoughts and other events. It seems clear then, that belief in determinism is also incompatible with belief that one has ultimate control. I now turn to the question of whether rational deliberation requires belief in ultimate control as well as belief in difference-making control.

3.3 Requirement of Belief in Ultimate Control for Deliberation, and How Such Belief Can Be Compromised

In chapter 2 I used an example involving a manipulating device to illustrate one way in which an agent's sense of control over her deliberative process could be compromised. I refer there to the agent's sense of a general loss of control over the process, but the example can serve to show more specifically how an agent's sense of losing *ultimate* control over a deliberative process would undermine her ability to deliberate. Here is the example.

Imagine a normal process of deliberation where options are adduced and weighed, and an intention is formed as a result. Now consider Stacy, who is about to engage in such a process but becomes aware that whenever she is about to begin to weigh reasons for and against each option, a chip that has been implanted in her brain activates and alters the strengths of her desires relevant to the weighing process, although after she begins to weigh reasons the chip becomes inactive. So, for instance, if the decision to be made concerns

a choice between ordering Chinese or Italian food for dinner, she knows that if she begins to try to find reasons for and against each alternative, the chip will activate and, say, strengthen her fondness of tomatoes and her dislike of reading through long menus.

I claimed that Stacy would be correct to believe that her reasoning is efficacious in this situation, but that her sense that her deliberation is under her control would be compromised to the extent that she would find it very difficult to form an intention about what to do with respect to this particular decision.

For present purposes it is unimportant whether the chip actually does function in this way. It does not even matter whether the chip exists, only that Stacy believes it does, and that she believes it sometimes works in this way. The point is that if she believes there is some reason to be uncertain that the strength of her desires relevant to the weighing process are really her own, in the sense that she, and nothing else, is their ultimate source, it seems very likely that she would in some sense *disown* her desires to the extent that she could not trust them to lead her to a decision that is really her decision – a decision that is an accurate expression of what she considers to be her true self. Realizing this, it's plausible that she would not be able to deliberate at all, as any desire that would ordinarily serve to help her evaluate her perceived options in a deliberative situation would be perceived by her as external, or detached from the process (in the sense described in chapter 2).

Interestingly, we have more than thought experiments to evaluate my claim that believing one is the ultimate source of one's thoughts is required for deliberation. In a phenomenon referred to as *thought insertion*, some schizophrenic patients claim to experience thoughts they did not produce. Although the 'inserted' thoughts sometimes have propositional form (I am worthless/stupid/evil; I am becoming a

robot; the government is spying on me), more often they take the form of commands believed to be issued by, for example, the government, evil beings, or aliens, that require patients to perform particular actions or tasks. The kinds of commands range from the relatively harmless (go to the store; stab myself in the arm with a pen), to the more dangerous (remove a bomb from my neck; kill my psychiatrist's family). The inserted thoughts are often experienced as voices, but sometimes simply as alien thoughts that are neither produced by, nor able to be controlled by the patient.

According to one patient: '... when I am thinking in this way, without being able to stop it... I have no mastery over the course of these ideas... it seems to me as if it is not me who generates them...' (Parnas and Arnorsson 2001, 108). Although there is no clear consensus about the nature or underlying causes of thought-insertion, a very appealing suggestion commonly found in the literature since Freud describes patients' experiences of lacking 'authorship' over their thoughts as a kind of withdrawal or distancing from those thoughts – an inability to include the thoughts in processes of practical reasoning. There appear to be two common strands of opinion on how best to characterize the phenomenon. As Christoph Hoerl (2001, 189) writes:

According to one view, patients' reports of thought insertion reflect an experience in which the boundary between the self and the world has dissolved, or at least an uncertainty, on the part of the patient, over how to draw this boundary. According to the other view, patients reporting thought insertion should be interpreted as articulating a distinction between thoughts that occur to them and thinking as something they do. Thus, what they mean is that they lack a sense of active participation in the occurrence of certain thoughts.

The distinguishing feature of this second view is the *passive* relation that an agent can seem to bear to her own thoughts: the thoughts ‘occur to’ her, rather than their being produced by her. It seems appropriate to describe this relation as one of externality or detachment, in the sense described above. The agent experiences herself as a mere spectator of her thoughts, rather than their author. So it is understandable that she might infer the source of these thoughts to be something other than her (aliens or evil beings, for instance).³ Hoerl (194) goes on to explain his own suggestion as to the likely causes of thought-insertion. Although somewhat more complex, it appears to be along similar lines.

There is a specific breakdown in schizophrenic patients’ reasoning abilities, which sometimes has the result, as we might put it, that the patients simply find certain issues settled. What they lack is a sense that certain thoughts, which come into their minds, are still theirs to consider, because the ability to take relevant alternatives into account is disrupted. If this is true, however, it may well be able to explain a sense in which the patients also do not experience themselves as having settled the issue. They are mere bystanders to the occurrence of those thoughts and thus may form the impression that those thoughts have been imposed on them by an outside agency. ... In short, the phenomenon of thought insertion arises because of a withdrawal, on the part of the patient, from certain forms of active engagement in reasoning.

Although some writers on the subject of thought-insertion (for example, John Campbell, 1999) have hypothesized that the phenomenon is best characterized as a

³ The inference I refer to here is almost certainly not as straightforward as I describe it. Mentally healthy agents also often seem to experience certain thoughts as simply occurring to them. Artistic and intellectual inspiration are perhaps examples of this kind of experience. However, there is at least one important difference between these two kinds of cases, relating to what I will call integration and ownership. I have more to say about this below.

breakdown in the mechanism that mediates the causation of occurrent thoughts by background beliefs, desires, and interests, this view by itself seems unable to accommodate cases in which thoughts described by patients as alien, or as not having originated with them, actually cohere well with their background attitudes and beliefs, and seem to have been produced by them. Due to the prevalence of cases like this, Hoerl reasons that for a thought to ‘belong to me’ it must not only cohere with my current beliefs and desires; it must also play a certain role in affecting the future background of beliefs, desires, and interests that I have, where this is determined by the active use I make of that thought in further deliberation. One patient describes his situation as one in which his ‘concerns’ have withdrawn to an ‘ideological domain.’ Hoerl thinks this kind of description is a good indication that ‘information is not gathered in a way that is informed by an active interest in making up one’s mind. What seems to be lacking is a certain form of self-determination, in which one uses one’s experiences and reasoning to form the shape of one’s future set of beliefs, desires, and interests.’ Similarly, ‘the point of identifying herself as the subject to whom certain occurrent thoughts belong may indeed be lost to someone who has withdrawn from activity to such an extent that her thinking is purely guided by what [Richard] Moran calls “theoretical questions”’ (2001, 199).

In order to try to clarify these suggestions, it seems that we can distinguish at least two ways in which the relation between a person’s background beliefs and desires and her current thoughts might play a role in her having an experience of an inserted thought. According to Campbell, a thought might be experienced as alien due to its failure to be produced by these background beliefs and desires in the appropriate way (perhaps in a way that is minimally coherent with them). It would follow that if a thought, such as a desire to kill, did not cohere with one’s background beliefs and

desires, one could not easily make sense of it as an expression of oneself, and so may have the impression that it had originated with an agency or source external to oneself. This hypothesis would be plausible in cases where the alien thoughts in fact do not cohere well with the subject's background beliefs and desires, or are clearly not produced by them. But often this does not seem to be the case. For instance, the subject may often have had the desire to kill, and may believe that killing is sometimes the right thing to do. She may even have a grudge against her potential victim. The second possibility incorporates the first, but includes what we might call a forward-looking component. The suggestion is that in order for a thought to be experienced by a subject as her own, not only must it be produced by her existing background beliefs and desires in an appropriate way; it must be recognized or entertained by its subject in such a way that it is able to produce, or play some part in producing, further beliefs and desires that cohere with existing ones – to 'form the shape of one's future set of beliefs, desires, and interests' as Hoerl puts it. This will be the case if that thought can be utilized in a process of deliberation – an intention-producing and belief-producing process. So rather than its being the case that the alien-thought-experience results merely from a breakdown in the mechanism that produces thoughts from existing background beliefs and desires, as in the first suggestion, this proposal highlights the possibility that a thought fail to be incorporated into a reasoning process that is capable of producing further beliefs and desires. The emphasis here is on what we might call the *ownership* of a thought rather than its authorship alone, and the idea points to a close connection between the experience of ownership of a thought and the experience of active participation in reasoning and deliberation. If this is right, it seems that at least part of the reason why a person experiences a thought as not belonging to her is that she stands in a passive relation to it. We might say that the

thought fails to be integrated into the subject's psyche. This distinction between authorship and ownership of thoughts might be characterized as follows.

AUTHORSHIP: The experience of authorship of a thought consists in the experience of the thought's being actively produced or initiated by the agent as an expression of her agency (this essentially captures the notion of being the ultimate source of one's thoughts, as described in the previous section).

Further, a thought one experiences having authored is likely to cohere with one's background beliefs and desires.

OWNERSHIP: The experience of ownership of a thought consists in the experience of the thought's being integrated with the agent's background beliefs and desires in such a way that she stands in an internal relation to it (as described in chapter 2). Such a thought will be capable of playing a part in mental processes that produce further thoughts, e.g., beliefs and desires.

I argued in chapter 2 that deliberation is essentially an active process that requires the agent to stand in an internal psychological relation to it. The phenomenon of thought-insertion, as well as the thought-experiment involving external manipulation of an agent's preferences, seem to indicate that unless one can assume that a thought has been authored by oneself, that thought cannot play a part in deliberation. Hoerl's view on the nature of thought-insertion seems to indicate a further point, which is that in order for a thought to be utilized in a process of deliberation, the thought must be integrated into the agent's psyche sufficiently for it to be capable of producing further beliefs and desires. In terms of the above characterizations, an agent must not only experience authorship of a thought, but also ownership of it.

Now, it is true that we often experience certain thoughts unexpectedly. On occasion we may even be surprised or shocked by the thoughts we have. For instance, a particularly violent image may ‘enter our heads’ as we are falling asleep, or we may hear an annoying and unwanted ‘little voice in our head’ which ‘tells us’ that we should do something we feel strongly inclined not to do.⁴ Surely we don’t always believe ourselves to have ultimate control over the production of these kinds of thoughts, and yet it may seem as though they could sometimes play a part in deliberation. So, is my claim that deliberation requires us to believe that we have ultimate control over our thoughts relevant to the deliberative process too strong?

In order to find out, we need to look more closely at the process by which an unexpected thought might be supposed to enter into a deliberative process. However, I should first emphasize a distinction made in chapter 2 between thoughts to which we bear an internal relation (those that are integrated into our psyche, forming part of what we experience as our self) and thoughts that are external to us, and which we experience as objects of our thinking. Part of my argument there implied that an agent must maintain an internal relation to the aspects of her psyche that are utilized for evaluating her options during a process of deliberation. For example, a desire to indulge in chocolate cake cannot be utilized in a deliberative process aimed at working out whether to attempt to satisfy that desire so long as that desire is itself being questioned or examined by the agent (as may happen, for example, if the agent starts to wonder whether her desire for chocolate cake has been prompted by her having recently seen an advertisement for chocolate cake – a type of factor she ordinarily tries

⁴ It is interesting to note that disorders such as Obsessive Compulsive Disorder have been characterized as the inability to rid oneself of unwanted thoughts. We could perhaps see some mental disorders, including the thought-insertion that can occur with schizophrenia, as extreme cases of what I am referring to as ‘unexpected thoughts.’

to ignore when deciding how to act).⁵ The agent must stand in an internal relation to the desire in order for it to be effective in helping her evaluate or decide something. If this is accurate, then it would seem that in order for an unexpected thought to be of any help to an agent in working out what to do, it would first have to be internalized or integrated. At the least this would involve its being made consistent with other relevant aspects of her psyche, for example, her attitudes, desires, interests, and preferences concerning chocolate cake, as well as perhaps food and eating more generally. In the simpler terminology used above, even if the agent does not initially experience authorship of a thought, she must experience ownership of it in order for it to play a part in her deliberation. And, as I explain below, with a sense of ownership is likely to come a sense of authorship.

There seem to be several possibilities with respect to the nature and outcome of the process of integrating an unexpected thought. One possibility is that the thought is rejected and forgotten, fails to be integrated, and plays no part in deliberation. Another is that the thought is modified in some way so that it can be integrated or owned. On a variation of the above example, the initial thought might be an unexpected desire for chocolate cake, triggered by the agent's noticing a chocolate cake on display at the supermarket. The desire is unexpected perhaps in part due to the agent's inability to remember previously having desired chocolate cake, and the fact that the desire does not cohere well with some of her other desires and beliefs (for instance, she does not

⁵ In some respects my claims here reflect the Stoic idea that agents have the ability to suspend judgment with regard to propositions such as 'It is fitting for me to eat chocolate cake,' which correspond to impressions such as the agent's noticing the picture of cake on the advertisement (cf. Long and Sedley 1987, 322; Pereboom 1994, 593). On the Stoic view, however, the agent's desire to eat the cake would not come about prior to the agent's having assented to the proposition that it is fitting for her to eat it. This view thus attributes to agents a degree of control over, and responsibility for the formation of their desires that is not present on the picture I have given, where the desire to eat chocolate cake is experienced directly as a result of the agent's recognizing the picture of the cake. Nevertheless, my notion of an agent's being detached from a desire due to her examining or questioning it is similar to the Stoic idea of an agent's suspending judgment of a proposition involving her eating cake.

want to put on weight, and believes that foods such as chocolate cake are high in calories which will contribute to weight gain). She thus treats the desire with suspicion, so it is not integrated at this point. Perhaps though, the agent considers this desire for a moment rather than simply dismissing it (by, say, entertaining her desire to not put on weight), and remembers that she had recently decided to experiment with tasting new foods and try to worry less about her weight. She now begins to endorse the desire for chocolate cake, as she finds it to be consistent with the more general desire to try new foods and worry less about her weight. So the desire for chocolate cake comes to be integrated, and she can utilize it in a decision-making process aimed at working out, e.g., whether her budget will allow her to buy some, how much to buy, whether she should buy the one she has noticed or a different one, etc. In a case like this though, it is plausible that what remains of the original unexpected desire once it has been integrated would be experienced by the agent as a thought that she has initiated – an idea that she, qua agent, has actively produced – as she has now found a way of justifying it and having it cohere with some of her background beliefs and desires. She may believe throughout this process that her desire for the cake was initially triggered externally. But due to her ability to distinguish between external stimuli that trigger desires in her and those that don't, she is unlikely to continue to regard the desire as alien due merely to its having been triggered by an external source. Upon reflection, if she continues to wonder why the desire initially seemed alien, she would be more likely to account for this by remembering that it initially seemed not to cohere with her other desires, rather than its having been triggered by the cake. Importantly though, it seems implausible that she would continue to regard the initial desire for chocolate cake as a desire that was hers but that she did not produce (a desire that she owns yet did not author), especially once the ongoing or

recurring desire for chocolate cake which had the initial desire as its basis, has played a part in further deliberative processes, and has thereby become more familiar to her.

Exactly how this process of integration works is unclear, but it is very likely similar to the case of artistic or intellectual inspiration. The inspirational idea may initially be experienced as unexpected and perhaps alien, but unless it is immediately rejected and forgotten, in time the inspired agent will come to regard the idea as something she owns, and further, something she produced. The more she works with it and becomes familiar with it, the more it will begin to cohere not only with further thoughts she is likely to have on its basis, but with her previous thoughts and background aspects of her psyche (beliefs, desires, etc.) – it will become an integrated part of her self. And once this process has taken root, as in the previous example, it seems unlikely that the agent would continue to regard the idea as something she did not produce, especially because her doing so would have to involve her similarly regarding any subsequent thoughts that had been built upon or derived from the original, as fully or partially alien.

Yet another possibility with respect to an unexpected thought is that the thought is not entirely rejected or forgotten, but remains to be considered at a later point. Here though, the agent will be in a position of uncertainty with respect to the thought until it is either integrated or rejected and forgotten, and it could therefore not plausibly function in a way that would help her to evaluate something else. There are perhaps other possibilities to consider, but in none of those discussed here does it seem that unexpected thoughts provide a clear counterexample to my claim that deliberation requires that one believe oneself to have ultimate control over one's thoughts relevant

to the deliberative process.⁶ Of course, we cannot ignore the possibility that an unexpected thought might be experienced as integrated and owned, yet not initiated by the agent. I have tried to show, with the above two examples, that this would be psychologically implausible. But I am unable to conclusively establish this. Nevertheless, this should not adversely affect my argument overall. Even if my contention were to be restricted to the claim that an agent needs to believe that she has ultimate control over her thoughts relevant to the deliberative process in the ownership sense alone, and not also in the authorship sense, event-causal libertarian accounts, as will be seen below, cannot clearly provide this.

This evidence from thought-insertion, as well as my claims about unexpected thoughts, are inconclusive and speculative. But it is important to reiterate the dialectical point made above, which is that, as argued in chapter 2, applying a belief in determinism to one's deliberative situation should result in one's inability to deliberate. Above I provided some support for the relatively uncontroversial claim that believing that one is causally determined is incompatible with believing that one is the ultimate source of one's thoughts and actions. The considerations raised in this section have been aimed at establishing the plausibility of the claim that belief that one is the ultimate source of one's thoughts relevant to a deliberative process is required for deliberation. If this claim is indeed plausible, then it would seem that finding grounds to think that we can be the ultimate source of our thoughts in a deliberative process, and thus that we can exercise ultimate control over them, is of the upmost importance.

⁶ I am not here claiming that an agent must *be* the ultimate source of a thought in order for it to be capable of playing a role in a deliberative process. If the authorship of a thought (a desire, for instance) is not questioned by its subject, and if the thought coheres minimally with her background beliefs, desires, etc., there seems no reason why the subject could not stand in an internal relation to the thought and utilize it in a deliberative process. My focus here is restricted to thoughts that strike an agent as alien to some extent – as perhaps not cohering particularly well with her current stream of consciousness or her background beliefs and desires, and thus being experienced to some extent as alien. My point is that once the agent becomes suspicious of the origin of a thought, she then stands in an external relation to it, and it cannot play a part in her deliberation until it is (re)integrated.

Once the question of control over our thoughts arises, we are not warranted in simply assuming that the falsity of determinism would be sufficient for the kind of control required, nor that *belief* in the falsity of determinism would be sufficient for *belief* that we have such control. What we require from an account of deliberation that assumes an indeterministic factor as an essential component is reason to think that the sense of ultimate control that is unavailable to agents who consider their deliberative situation to be causally determined would be attainable due to this indeterministic factor. We need a good explanation of *how* indeterminism could be of any help. For example, the libertarian assumption that there are alternatives to the events that actually occur does not warrant the further claim that we can *control* which of those alternatives becomes actual. And (more importantly for present purposes) given the assumption that causal chains may begin in our brains, we are not warranted to make the further claim that they are produced *by us* or that we are in control of their initiation. I will use Robert Kane's account as an example, due to its comprehensiveness and detail with respect to the points that will primarily concern me. In the concluding section of the chapter I will relate my findings to other non-agent-causal libertarian accounts.

3.4 The Basics of Kane's Account of Deliberation

According to Kane (1996), a process of deliberation may be influenced by indeterministic occurrences that would be beneficial to the deliberating agent. Far from undermining the rationality of the agent, or his control over the process, indeterminism can play a creative role (164), 'opening doors' in deliberation, and making possible 'new beginnings' (165).

If undetermined occurrences play a role in deliberation leading to practical choice, they do so by initiating processes of thinking about, or imagining, remembering, or attending to various facts, memories, images, and scenarios that may be relevant to deliberation. We refer generally to the items thus thought about, remembered, or attended to as ‘chance-selected considerations.’ Some of them are passed over as irrelevant to deliberation, but others are acknowledged by the agent as relevant. This acknowledgment of relevance amounts to the adoption into the preference set for one or another option of reasons not previously regarded as relevant to the deliberation, or not previously seen as relevant in this way (p. 163).

Kane thinks it’s plausible that these chance-selected considerations ‘may well up from the unconscious mind,’ influencing the stream of consciousness, ‘suggesting new options, new consequences of the options, and new ways of viewing the consequences’ (160). Although these considerations are ‘chance-selected,’ and thus we are not able to control which of them are going to enter our consciousness, they are not completely beyond our control. By what Kane refers to as a ‘Taoist effort,’ we can relax our mind, opening ourselves up to new thoughts, thereby allowing ourselves to be receptive to new chance-selected considerations (165). Kane specifies that these Taoist efforts require us to resist our natural tendencies to think in familiar ways, and so they are made against resistance. So the efforts are reflected in the brain by movement further from thermodynamic equilibrium, and this makes the process sensitive to undetermined influences on the micro-level (166). These micro-level undetermined influences (we can assume from what he says earlier) are then amplified chaotically, resulting in the macro-level phenomena that are the introduction of new considerations into the deliberative process. So deliberators can do more than just passively wait for these chance-selected considerations to occur. We can make ‘efforts

to temporarily relinquish conscious control over thought processes in order to be receptive to new considerations that may come to mind – that is, efforts-not-to-make-an-effort to control [our] thoughts’ (165).

These Taoist efforts are not the only kind of indeterministic effort that can have a positive influence on deliberation. Kane thinks that agents can also make indeterministic efforts of the following kinds: ‘(1) efforts not to quit deliberating too soon or not to choose too hastily, (2) efforts to focus attention on considerations that have come to mind in order to determine all their consequences, (3) to avoid suppressing relevant considerations that one may be resistant to considering, (4) to galvanize oneself to pursue relevant sources of new information, and (5) to avoid deceiving oneself (or rationalizing about) the relevance of various considerations or their consequences’ (167). I address these other kinds of effort briefly in section 5b and in more detail in section 5e. Of particular significance for our present concern, Kane points out that it is through our making such efforts that ‘we exercise control over our practical deliberations throughout their duration’ (168). This is the point that I will be contesting.

3.5 Antecedent-Determining Control and Ultimate Control

Before assessing Kane’s account for its ability to secure the control required for deliberation it will be useful to try to clarify the distinction he draws between ‘antecedent-determining’ control and ultimate control. He describes antecedent-determining (AD) control as ‘the ability to be in, or bring about, conditions such that one can guarantee or determine which of a set of outcomes is going to occur *before* it occurs, whether the outcomes are one’s own actions, the actions of others, or events in the world generally’ (1996, 144). As this kind of control is compatible with the truth

of determinism, it cannot by itself be exhaustive of the control required of a libertarian account of deliberation. Kane's notion of ultimate control is less clear. Here are the two most illuminating things he says about it: '...it does not follow that because you cannot determine which of a set of outcomes occurs *before* it occurs, you lack control over which of them occurs, *when* it occurs. When the conditions of plural voluntary control are satisfied, agents exercise control over their future lives *then and there* in a manner that is not antecedently determined by their pasts' (144). Similarly, 'Though you can't guarantee in advance for any one of the options taken separately that it will occur, you have complete control over the options considered as a set, in that you can do whichever of them you will to do *when* you will to do it, for the reasons you will to do it, on purpose, without being coerced, compelled, etc.' (1999, 117). Plural voluntary control is defined as follows: 'Agents have plural voluntary control over a set of options (e.g., choosing morally or prudentially, or vice versa) when they are able to do whichever of the options they will to do, when they will to do it..., for the reasons they will to do it..., on purpose rather than accidentally, by mistake, or merely by chance..., without being coerced or compelled in doing it... or in willing to do it..., or otherwise controlled in doing or willing to do it by any other agents or mechanisms' (1996, 143).

Part of the difficulty in assessing Kane's account of deliberation stems from the difficulty in understanding what kind of control he has in mind here. As it is an indeterministic account, we must assume that it is an element of indeterminism that gives this kind of control a particular significance in his overall view and that distinguishes it from the kind of control that we could have if determinism is true (AD control). In order to identify this distinctive indeterministic element we must isolate it from the components in these definitions that are compatible with determinism. As I

see it, the only distinctively indeterministic aspect of Kane's notion of ultimate control is the idea that you have control over which member of a set of outcomes occurs *when* it occurs (rather than *before* it occurs) in a manner that is not antecedently determined by your past. But it is not clear exactly what this amounts to. It is important to remember that Kane's account is *not* agent-causal, whereas the notion of being the ultimate source of one's actions, as described in section 1 above, seems to be most naturally interpreted as involving causation between the agent and her thought in a way that is not clearly reducible to causation between mental events. Therefore, any descriptions given by Kane involving agent-causal language must be treated with suspicion, especially if he is using them to explain a causal connection that would be mysterious without the assumption of agent-causation. For example, Kane claims that '[a] free action for which the agent is ultimately responsible is such that its occurring rather than not here and now, or vice versa, has as its ultimate or final explanation the fact that it is *caused by the agent* here and now' (1995, 120 (my emphasis)). But it is illegitimate for descriptions involving causation by agents to play any irreducible explanatory role in accounts that reject the agent-causal relation, as Kane's explicitly does (cf. 1996, 115-17). Therefore, such descriptions must be either treated as shorthand for reductive explanations purely involving event-causation, or else rejected.

My strategy will be to assess Kane's account in terms of whether it plausibly captures the idea that an agent can be the ultimate source of her thoughts in a way that can be psychologically internalized by the agent and allow her to stand in an internal psychological relation to them.

3.6 The Search for Ultimate Control on Kane's Account

We now need to look for some aspect of Kane's account in virtue of which an agent could believe that she is the ultimate source of her thoughts,⁷ i.e., that she has ultimate control over them. As we have seen, and as will become more apparent below, Kane is not altogether clear about precisely what he intends to denote by 'ultimate control.' In order to account for any terminological discrepancies, therefore, I will reserve the capitalized version ('Ultimate Control') for the kind of control we are looking for, i.e., the kind of control that an agent needs to believe she has in order to deliberate. As a further note about my approach, the following is intended to be a conceptual analysis of Kane's account of the process of deliberation, in order to pinpoint where control would be relevant. Deliberation often occurs very quickly, and the phases described below may overlap one another to some extent. But I think it's legitimate to break things down in this way, as it seems the most promising method of obtaining an explicit picture of just what kind and degree of control could be involved.

3.6a The Irrelevance of AD Control

Since I've already argued in chapter 2 that belief in determinism is incompatible with belief that one has the kind of control required for deliberation, that is, that one is the ultimate source of one's thoughts, I won't focus much attention on parts of the deliberative process on Kane's account where it is clear that the only control available to the agent is AD control. For example, although he does not make it explicit, it is implied by some of what Kane says that deliberators, on his account,

⁷ I have generally been referring above to the requirement of ultimate control over one's thoughts *and* actions. Henceforth I will be restricting my focus to an agent's thoughts, until I look at the possibility of an agent's exercising ultimate control over her intention-formation, in section 3.7. In doing this I do not intend to exclude mental actions such as willing and trying as potential sources of ultimate control.

would sometimes be adducing and weighing reasons deterministically, that is, there would be no indeterminism involved in that part of the process. We know that on his account of the Taoist effort, indeterminism only plays a part in the process when agents make efforts to resist familiar modes of thought. This resistance creates tension in the will, giving rise to indeterminacies that are amplified chaotically and amount to the coming to mind of new ‘chance-selected’ considerations (CSCs). But he claims that deliberation can proceed without the Taoist effort’s being made: ‘The point is not that Taoist efforts [to resist normal ways of thinking and relax the mind] must be made or always are made in practical deliberation’ (1996, 165). Kane also claims that we *relinquish* conscious control over thought processes when we make Taoist efforts (165). This implies that we can be in control of the process up to the point at which we relinquish it by making the Taoist effort. But notice that whatever kind of control we have *before* the point at which we make the Taoist effort, it could not be ultimate control, as the process up to this point involves no effort to resist normal ways of thinking, and therefore there is no tension in the will which would give rise to the indeterminacies that must play a part in our exercising ultimate control. Of course, we may *feel* as though we are in control of this part of the process, but phenomenological evidence such as this cannot legitimately be used to establish that we are in fact in control. An incompatibilist such as Kane should certainly not give such evidence much weight if it were being used to establish, for example, that we can be in control of our thoughts and actions if determinism is true. Due to these considerations I think we can safely restrict our search for Ultimate Control on Kane’s account to those parts of a deliberative process that are likely to include indeterministic occurrences.

3.6b Initiation of the Taoist Effort

We now need to turn our attention to parts of the deliberative process where Ultimate Control is likely to be found. As the Taoist effort seems the obvious candidate, we should start there. Kane describes the Taoist effort in general terms as an effort made by a deliberator to relax her mind, with the intention of finding new considerations relevant to her decision (1996, 165). Although Kane does not do so, I think a distinction must be made between the Taoist effort itself and its *initiation*, i.e., the precise point in time at which the effort begins. This will allow us to more accurately assess the account for Ultimate Control. Now, there is some evidence that Kane thinks an agent is in control of the initiation of the Taoist effort. He says that voluntary efforts are under our control, or at least that the indeterminacy involved in voluntary efforts is relevant to control and responsibility (1999, 117). Also, he claims that we consciously intend to make the effort *before* it is made, so it looks as though this may involve at least AD control, provided that the intention causally necessitates the effort's being made.

So far this is quite general, merely providing some evidence that Kane thinks we can control the initiation of the Taoist effort in some respect. But he goes into more detail in a response to Clarke, who asks how the agent controls *whether* an effort is made. Kane (1999, 116, n. 10) claims, first, that it's not the case that the effort is something someone controls by doing something else. Rather, the effort is 'the agent's doing something in order to control something else (e.g., which choice is made or whether an overt action is performed).' The agent does cause her effort, but this means only that she makes it, not that she causes it by doing something else (which he thinks would lead to a regress). He goes on to explain that on the neural level this effort-making 'would be a complex neural process involving circulating impulses in a

recurrent neural network.’ This process, in turn, is caused by the neural processes that realize the agent’s reasons for making the effort (her beliefs, desires, etc.) These comprise the agent’s motivation for making the effort – they cause the effort to be made. He is careful to remind us, however, that these reasons do not determine the effort, because the effort is indeterminate, meaning that its effects are undetermined, and also that its onset (whether it occurs) is undetermined.⁸ Further, the reasons do not determine the choice that results from the effort, because it is undetermined whether the effort will succeed.

We need to look at this more carefully. Let’s start by assuming that the agent has a reason for making the effort, e.g., a desire to have new considerations relevant to her decision occur to her, and a belief that it may help if she tries to relax control over her thoughts. About a reason such as this Kane says both that it motivates her to make

⁸ Kane explains what he means by ‘indeterminate,’ as well as the distinction between ‘indeterminate’ and ‘undetermined’ by drawing on the postulated indeterminacy of properties of physical systems (e.g., the position and momentum of particles) by quantum physicists. He uses an analogy to illustrate his idea that an effort of will could be indeterminate: ‘Imagine an isolated particle, such as an electron, moving toward a thin atomic barrier. Whether or not the particle will penetrate the barrier is undetermined. There is a probability that it will penetrate, but not a certainty, because its position and momentum are not both determinate as it moves toward the barrier’ (1996, 128, 172-3). To help explain this further, Kane distinguishes two kinds of indeterministic worlds. In what he calls an ‘Epicurean’ world ‘all properties of physical systems, such as position and momentum, are determinate at all times, but... (unlike the classical Newtonian world, as usually envisioned) the laws allow different (determinate) futures given the same (determinate) past.’ This is contrasted with a ‘non-Epicurean’ world, which is ‘like the quantum world of modern physics (on standard interpretations of it) that allows for both indeterminateness of physical properties and the possibility of forks in history’ (172-3). So for an effort of will to be indeterminate is for it to correspond to a complex macro process in our brain, involving many neuron firings and connections which have indeterminate physical properties on the quantum level. These quantum indeterminacies are amplified chaotically such that the strength of the effort is indeterminate, and therefore its effects are undetermined. A potential source of confusion here is Kane’s reference to indeterminate processes or events, e.g., ‘indeterminate acquiring or possessing of reasons’ (1999, 117). It is unclear why Kane refers to these events as indeterminate rather than undetermined, given that other uses of ‘indeterminate’ refer to properties such as the position and momentum of a particle and the strength of an effort, where there is a reasonably clear analogy to quantum physics. But whatever precisely he means by this, I think it is consistent with the quantum analogy given above, and will not adversely affect my analysis below, if we assume that an indeterminate event or process is one that is indeterministically caused, i.e., undetermined by its causes. Surely he does not mean by ‘indeterminate’ in this context that after an agent has acquired a reason there is merely a probability that, i.e., no fact of the matter whether, the agent now possesses that reason.

the effort and that it causes her to make it. Now let's look at the precise point at which the effort is initiated. Again, Kane claims that the effort is indeterminate, which for present purposes we are understanding as its having a probability less than 1.0 of occurring given what would be its causes: '[The] reasons or motives which causally influence the effort do not *determine* it exactly because it is not *exactly* anything; it is an indeterminate effort' (1999, 16, n. 10). He also says that the (initiation of the) effort is caused by the agent, but not by her doing something else.

Assuming this is coherent, we need to know whether the agent *controls* the initiation of the effort. One possibility is that she doesn't. If this is the case, then we can move on to consider whether she controls the effort post-initiation. There is some evidence that Kane would agree that the agent does not control the initiation of the effort, when he says, for example, as mentioned above, that Clarke's question of how the agent controls whether an effort is made is misleading if it presupposes that the effort is something someone controls by doing something else, rather than being the agent's *doing something in order to control something else*. The possibility that control enters the process at a later stage rather than at the point the effort is initiated is discussed below in section 6c. But to be safe, for now let's proceed on the assumption that Kane thinks the agent *does* control the initiation of the effort. Now given that 1) her reasons cause her to make it, 2) she causes it, 3) she does not cause it by doing something else, *and* given the further assumption that 4) she does not agent-cause it, we have to assume that there is a direct causal link, albeit an indeterministic one, between her reasons for making the effort and the effort's being made. So given that there is no agent-causation in play, the expression 'she causes it' is presumably to be understood to mean that her reasons cause it, and plausibly that she is consciously aware of causing it. But this raises the question 'How might the agent be in control of

her reasons causing her effort, or whether her reasons cause her effort?’ After all, we can certainly have reasons for doing something that do *not* cause us to do it and that do not cause it to occur. So what is the extra factor that makes it the case that her reasons change from being dormant to being active and initiating the effort? It can’t be that the agent performs some mental action that, in turn, causes her reasons to cause the effort, as this is ruled out by claim 3. So it would seem that the only possibility remaining as to how her reasons become active and cause the effort to be made is to be found in the indeterminism. Unless we have some other way of understanding claim 2 in conjunction with claims 3 and 4, the indeterminism involved with claim 1 is the only factor remaining which is a potential explanatory ground for the agent’s control over her initiating the effort. But indeterminism is not an extra factor that supplies a cause; ‘indeterministic’ is merely a term used to describe a type of causal link where the cause doesn’t necessitate the effect. Certainly we can say that a reason became an active cause indeterministically, i.e., it was dormant and then indeterministically became active and caused the initiation of effort, with nothing in turn, or concurrently, causing this activity to take place at that point. But this supposition will surely do nothing to enhance the agent’s control. That is, without some extra factor in play it is implausible to suppose that the agent controlled *this*, or that this constitutes the agent’s control.

My point here can be brought out more clearly if we look at something Kane says about our having control over indeterministic occurrences. He says that ‘the important point for control and responsibility would again be whether these indeterminate acquirings or possessings of reasons just “happen” or “occur” spontaneously to the agents or whether they result in part from the agents’ *voluntary* efforts or *doings*’ (1999, 117). This passage is directed at Clarke’s suggestion that the

agent's acquiring reasons might be indeterministic processes. But the contrast Kane makes here is illuminating nonetheless. It is illuminating because in using it he seems to be implying that indeterministic occurrences can be under the control of agents by virtue of their resulting from voluntary efforts of agents. But this strategy is inadequate to establish control unless we assume that the phenomenology involved with a voluntary effort is a reliable indicator of actual control. Now if we *do* assume this, then it clearly does not enhance the account, for as mentioned above, it is illegitimate for Kane to rely on phenomenological evidence in this way. But if we don't take such evidence on face-value then we need an explanation of how the initiation of an effort that results in indeterministic occurrences is in turn under the control of the agent. And this is precisely what we don't have.

Of relevance here, Clarke (2003, 89) says that 'if efforts preceding decisions in cases of moral conflict are supposed to contribute in this way to active control, then what is needed is an account of the freedom with which the agent acts in making these efforts.' Clarke claims (2003, 90) that Kane has responded (in correspondence) that the efforts preceding decisions are not directly free (given incompatibilism), i.e., they are free only in a compatibilist sense (uncoerced, uncompelled, etc.) But, according to Clarke, 'given the assumption of incompatibilism, we cannot get freedom-level active control for the decision by adding together the compatibilist "freedom" of the prior effort with whatever direct active control is exercised in making the subsequent decision, unless the latter is already (by itself) freedom-level active control... Freedom-level active control cannot result by addition in this way.'⁹ So we lack an adequate control-relevant explanation, in purely event-causal terms, of how the effort

⁹ And as Clarke points out (90 n. 28), Kane's notion of doubling, which refers to the agent's trying to do two incompatible things at once (see, for example, Kane 1999, pp. 111-14), would not be of any help.

is initiated. Aside from phenomenological evidence, there seems no reason to assume that our doing something amounts to our controlling our doing it.

In terms of *ultimate* control in particular, there may be another explanation available to Kane here, which is to concede that the agent has very little AD control over the initiation of the effort, but to claim that she does control it in the sense that she has ultimate control over it. He concedes that indeterminism diminishes AD control, but thinks that this is the cost of our exercising ultimate control. So we need to take another look at what Kane says about ultimate control and work out whether it applies to this part of the process. As mentioned above, Kane gives very little detail about the nature of ultimate control. We can at least be sure, however, that he thinks it applies at the point at which an agent chooses one out of a range of perceived options. Ultimate control is exercised when we choose which member of a set of outcomes occurs as it occurs, even though we can't guarantee which will occur before it occurs: 'Though you have diminished antecedent control over *each* of the options taken separately (you cannot guarantee in advance for any one of them that it will occur), you have complete or "plural" voluntary control over the available options at a given time considered as a *set* in the sense that you can do *whichever* of them you will to do, *when* you will to do it, for the reasons you will to do it, on purpose, without being coerced or compelled' (1999, 117 n. 13).

In explaining what kind of determining power ultimate control is, so that it doesn't look entirely negative (just the absence of determinism), Kane explains two notions that he thinks play a positive role in that account.

1. When there is conflict in the will due to an agent's desire to do two incompatible things, there will be resistance in the will to whichever is

chosen, so effort is needed to resist and overcome the opposing desire (115-16).

2. If the choice or action is the end result of a trajectory of the agent's prior effort, then the agent is responsible despite its being undetermined whether the effort would succeed (116).

It's unclear how these descriptions help to clarify the notion of ultimate control in a positive way. Ultimate control must involve something more than a combination of AD control, sense of control, and its simply being undetermined which of the competing reasons will win out. That further thing, let's assume, is the agent's prior (prior to one of the reasons winning out) effort to have a reason prevail. On this assumption, the agent exercises ultimate control over the prevailing of that reason because that was what she was trying to do. But we may wonder why it would be that these efforts are instances of ultimate control rather than AD control. On the evidence we have from Kane, it appears that there is no indeterminism in play at least up until it occurs to the agent that making a Taoist effort is a good idea. So in order for ultimate control to apply to this part of the process, we would have to suppose that beginning at the point at which this occurs to the agent there is tension in her will, because she both wants to make the Taoist effort and also does not want to make it. We must bear in mind that the tension must be in place *before* the effort is initiated in order for the agent to have ultimate control over its initiation.¹⁰

In the first place, it's questionable whether this is what would be occurring to a deliberator at this point in the process. In order to exercise ultimate control over the

¹⁰ It is important to point out that this 'doubling' of effort plays a central role in Kane's account of moral and prudential decisions, where it has more intuitive appeal. Here I am taking the (small) liberty of applying the notion explicitly to the case of deliberation in an attempt to make sense of ultimate control in this context.

initiation of the Taoist effort, the agent would first need to be consciously attending to her reasons for relaxing control, and also (simultaneously) consciously attending to her reasons to go on deliberating in a familiar way. Ultimate control on Kane's account is in place only when an agent is simultaneously trying to make two incommensurable sets of reasons win out and cause a decision, as in moral and prudential choice situations (see 1999, ch. 8). So in order for ultimate control to be in place at this point in the deliberative process, the agent must be simultaneously trying to make an effort to relax conscious control over her thoughts *and* to not relax control and just go on deliberating without the Taoist effort. But first, the idea of *trying to make an effort* seems problematic on a psychological level (and may involve an unhelpful regress). If we add to this that the agent is not only trying to make an effort to relax, but also trying to make an effort not to relax, I think we end up with an unacceptable tangle of tryings and efforts. And again, to square this with other parts of Kane's account where ultimate control is relevant to effort-making (e.g., moral and prudential choice), the agent would have to be trying both to make the effort (trying to make her reasons for making the effort prevail) and not make the effort (trying to make her reasons for not making the effort prevail) in order to make a decision to make the effort. But again, this seems unacceptable, and given this reasoning, a regress looms. Since an exercise of ultimate control always requires there to be *prior* tension in the agent's will due to her trying to make competing sets of reasons prevail, in order for the agent to exercise ultimate control over the initiation of any particular effort it seems that we would need to keep positing further prior (competing) efforts in which the required tension could be found. As a final point, Kane says that 'since in practical choice situations agents are not making efforts to do what they think they ought to do against prevailing inclinations, such efforts cannot play the same role in practical choice as they do in moral and prudential choice – or, for that matter, in the

efforts sustaining purposes... . As a consequence, indeterminism cannot enter into practical choice in the same way it enters into other categories of SFWs' (1996, 159). From all of this it seems reasonable to conclude that agents could not exercise ultimate control over the initiation of the Taoist effort.

So, according to this analysis of Kane's account, we possess reduced AD control over the initiation of the Taoist effort (due to the indeterminism), and we have no ultimate control over it. So we can conclude that we cannot have Ultimate Control over the initiation of the Taoist effort. We can therefore assume that the agent has less control over this point in the deliberative process on Kane's account than she would if it were purely deterministic. This analysis also applies to the other indeterministic efforts mentioned by Kane, and given in section 3 above. These are: '(1) efforts not to quit deliberating too soon or not to choose too hastily, (2) efforts to focus attention on considerations that have come to mind in order to determine all their consequences, (3) to avoid suppressing relevant considerations that one may be resistant to considering, (4) to galvanize oneself to pursue relevant sources of new information, and (5) to avoid deceiving oneself (or rationalizing about) the relevance of various considerations or their consequences' (1996, 167). Kane gives examples of these other prudential efforts, but does not go into any more detail about how the indeterminism enters into the deliberative process via such efforts, than he gives with respect to the Taoist effort. So without more information we can also conclude, for the reasons given in this section, that agents would have diminished AD control, and no Ultimate Control, over the initiation of these other kinds of efforts. (I address these other kinds of efforts again in section 3.6e.)

3.6c *The Taoist Effort and CSCs*

Even if an agent lacks Ultimate Control over the initiation of the Taoist effort, might she not have control in the post-initiation phase? Here, according to Kane, the agent is in the process of making an effort to relax her mind, which, because it goes against familiar ways of thinking (the deterministic adducing and weighing of reasons described in 6a above), creates tension in the mind which stirs up indeterminacies. These indeterminacies are amplified chaotically and amount to the coming to mind – from unconscious to conscious – of new chance-selected considerations (CSCs). So the question to be addressed in this section is whether an agent can have Ultimate Control in this phase, in particular, over the coming to mind of the CSCs.

One immediate difficulty in assessing this aspect of Kane's account is that he suggests in some places that we are in control of the entire Taoist effort and in others that we lack control of at least some parts of it. On the control side, he says, 'To the extent that we succeed or fail in our efforts to make such [informed, prudent] decisions, we exercise control over practical deliberations throughout their duration' (1996, 168).¹¹ Yet he also says that agents would either lack control or have diminished control over at least some parts of this phase of the process: 'By definition, reflecting agents cannot control exactly which "chance-selected" considerations are going to come to mind at any moment. But they can willfully put themselves in a frame of mind that is *receptive* to new chance-selected considerations' (165). Another (potential) problem is ascertaining just where in the process the indeterminism plays a role. For instance, he says that 'the indeterministic effort of will corresponds to the

¹¹ Additionally, he claims that 'these additional ideas [the efforts, including the Taoist effort]... involve giving agents a more active role in practical deliberation by way of efforts of will through which the agents might exercise greater control over the deliberative process – *without* eliminating the creative role of chance-selected considerations' (164).

entire chaotic process in the brain, including the micro indeterminacies and their amplifications.’ In order to explain this, he says, more specifically, that while the effort is being made, ‘the micro indeterminacies are being fed upward to the neural net as a whole, which is continually reorganizing in response to micro indeterminacies and is in turn influencing individual neurons.’ So, he claims, ‘the indeterminism and the effort are “fused”’ (151). But it’s hard to see exactly what this fusion amounts to, and how to distinguish the specific parts of the process that are undetermined from those that are determined. Kane does concede that the indeterminate efforts are mysterious, but claims that this is due to the difficulties of understanding consciousness and quantum indeterminacy.

Perhaps we can clear this up if we imagine some of these parts of the process occurring simultaneously. We could be in (AD) control of making the effort, which is an example of an effort of will sustaining a purpose (1996, 166), and during the effort-making phase (what we experience as making an effort) CSCs indeterministically occur to us as a result of the tension created by our effort, and yet we’re not in control of which and when they do (1999, 116). But we go on making the effort while this is happening. So in a sense we’re in control of the effort all the way through, but only in one direction, i.e., ‘towards’ the indeterministically produced CSCs. This is probably why Kane claims that the effort and the indeterminism are ‘fused’ – the indeterminism is a property of the effort, so the effort is indeterminate (1996, 151). So we could be simultaneously in control of something (our making the effort) and not in control of something else, the direct results of that effort in the form of the occurrence of CSCs. We may have the *experience* of making an effort throughout this part of the process, with considerations coming to mind while the effort is being made, like sporadic effects of a continuous cause. It seems that this is the only way to reconcile Kane’s

claims that on the one hand, the effort includes the indeterminacies (1996, 151), and we are in control during the entire effort (1996, 168), and on the other hand, that reason relaxes control (to allow for chance-selected considerations)...’ (165), and agents have no control over when and which CSCs will arise (1999, 116).

To collect results, it seems as though an agent would have very little, if any, AD control during this phase. According to Kane, agents have no control over when and which CSCs occur, and phenomenologically, an agent would have relinquished (or at least relaxed) conscious control over this part of the process in order to put herself in a frame of mind whereby she is receptive to the occurrence of CSCs. In terms of ultimate control, certainly an agent doesn’t have *any* kind of control over which and when CSCs occur. But she may feel like she is in control of making the effort to relax her mind and freely associate (the Taoist effort). As long as it makes sense to be making an effort to not be making an effort, and *succeeding* in not making an effort (while simultaneously making the effort to do this), this picture could be plausible. But there still may be a problem. For one thing, it seems fairly clear that whatever exactly Kane means by ‘ultimate control,’ given his aims it would have to apply at the point at which one of the agent’s competing sets of reasons prevails and leads to a choice being made (or in the case of efforts sustaining purposes, the prevailing of an intention over temptation to do otherwise). Now Kane does say that efforts of will cannot play the same role in deliberation as they do in moral and prudential choices or in efforts of will sustaining purposes, but without more detail about the role they do play in deliberation we will have to use their role in these other categories as a guide.

The picture must be something like the following. When making-the-Taoist-effort an agent has two incommensurable sets of reasons: one for making the effort to

relax her mind and one for not making it, and she is trying simultaneously to make both sets prevail. This tension causes the process to be sensitive to indeterminism so that the strengths of the efforts to make each set prevail are indeterminate, and which set of reasons prevails is undetermined. When one set does prevail (e.g., the reasons for making the effort to relax her mind) the agent exercises ultimate control *at that point*, as this is the end result of the trajectory of a prior effort of the agent. In moral and prudential decisions we must assume that after the point at which one of the sets of reasons prevails, there is no longer any tension and the process becomes deterministic once again. Now, due to the Taoist effort's being a 'frame of mind' like a meditative state (1996, 165), rather than a decision (or, more precisely, the formation of an intention, as in the moral and prudential case), in order for CSCs to occur to the agent throughout this phase, the agent must continually be in tension as a result of her effort to make both of her reason sets prevail. But if she exercises ultimate control during the entire phase, and ultimate control is only in place when one of her reason sets prevails (it would be control over that set's prevailing), it must be the case that she is continually trying to make both of her reason sets prevail *and* succeeding at making one of them prevail. Perhaps this is why Kane says that the effort-making 'would be a complex neural process involving circulating impulses in a recurrent neural network' (1999, 166 n. 10). Here's the point, though: According to Kane, ultimate control is exercised only at the point at which one of the sets of reasons prevails. And once one of the sets has prevailed the agent can no longer be in tension trying to make them both prevail. But CSCs occur only when the agent is in tension, as it is the tension which gives rise to the indeterminism – the element of chance which is an essential to CSCs. But while the agent is in tension she is not exercising ultimate control. So it seems that, to be very specific, the agent is only able to exercise ultimate control at the precise point when no CSCs could be occurring. So what exactly is she exercising

ultimate control over? The only possibility is that she exercises it over her success at relaxing her mind. And this would clearly be of no help to an agent who requires good grounds to believe that she has Ultimate Control over her thoughts during a process of deliberation.

3.6d Regaining Control

If we left the analysis here, we would have completed our search for Ultimate Control in the parts of the deliberative process where Kane indicates it is most likely to be found, i.e., the Taoist effort and the occurrence of CSCs. But we should also look at the possibility that Ultimate Control is in play in other parts of the process. It is important to note that Kane does not say much about parts of the deliberative process *after* CSCs have occurred to the agent.¹² So the accuracy of the analysis given in this section will depend on whether some of my previous assumptions about Kane's account are correct. Importantly, I'll continue to assume that at least some of the phases of deliberation can occur simultaneously. Even on this assumption, though, it will be necessary to identify and isolate the different causal and phenomenological

¹² In his chapter 8 (1996) Kane does describe stages in decision-making processes where he thinks an agent exercises ultimate control over her choosing between perceived alternatives for action. It may be thought that such descriptions could be helpful in illuminating the parts of a deliberative process with which we are presently concerned, i.e., after the Taoist effort has been made and the CSCs have occurred. This does not seem to be the case, however, as the kinds of decisions discussed in chapter 8 (moral and prudential) essentially involve a conflict between what the agent believes she ought to do and what she desires to do (126). It is this conflict or tension in the agent's will, and the agent's struggle to resolve it, that magnify the indeterminacies that play the essential role in allowing an agent to exercise ultimate control over her decision. But the important point is that the process is completed once this conflict is over and the agent has decided to either do what she believes she ought to do, or give in to temptation (i.e., 'set her will' one way or the other). By contrast, in 'practical choice,' or what I am referring to as 'deliberation,' the tension is caused by the agent's competing sets of reasons for making the Taoist effort or not making it (or by her competing reasons for making the other efforts involved in deliberation, e.g., efforts not to quit deliberating too soon). But the deliberative process does not end once one of these sets of reasons prevails. Hence, the descriptions given in chapter 8 of how the agent can exercise ultimate control are of little help here. Kane does mention, and briefly illustrate, five kinds of indeterministic efforts other than the Taoist effort, which he claims can enhance the agent's control over her deliberation. I discuss these in section 3.6e.

strands in the process, in order to avoid any unnecessary confusion that may result from the complexity.

So let's assume that the agent has tried to relax her mind in an effort to have new considerations occur to her. The tension this has created due to the attempt to relax being made against resistance has stirred up indeterminacies and new considerations have indeterministically come to mind from the unconscious. We'll take the simplest case first. This is basically linear. At this point the agent, for example, experiences a memory or imagines a scenario which had not occurred to her previously. It seems plausible, and consistent with Kane's account, that at this point the agent is no longer making a Taoist effort to relax her mind in order to be receptive to new considerations, as she now has something new and significant relevant to her deliberative situation to ponder in the stage that now begins, i.e., the weighing stage. I'm assuming for this case that the Taoist effort requires her whole stream of consciousness, so she doesn't both keep making the Taoist effort *and* deliberate about the new consideration with respect to her decision. So now she's back to deliberating deterministically (i.e., the causation in this phase is deterministic only). She regains full AD control. Why 'full'? Because during the 'fishing' stage of the effort, according to Kane, when she is receptive to new CSCs occurring to her, she had relinquished full AD control, but not all control. So we can assume that she had AD control over her making the effort, but not over what thoughts occur to her as she is making it. But, to look closely, she did not consciously will to regain full AD control; it was the occurrence of the new consideration that caused her to cease making the Taoist effort and begin pondering the new consideration, and this she did not control specifically. So we can fairly safely conclude that she did not control the initiation of her returning to deterministic deliberation.

Now let's look at a more complex possibility. Again, the agent makes the Taoist effort, and a previously unattended-to consideration becomes conscious. But she doesn't recognize it to be particularly relevant, so she continues to make the effort (which, we must remember, is an *effort*, and so requires concentration). So what happens to the new consideration? Let's say she keeps it in short-term memory. But there's no indeterminism involved with this, as there's no 'against-the-grain' effort being exerted with respect to this particular consideration.¹³ Once she's accumulated a few of these new considerations she wants to take stock, so she ceases making the Taoist effort. What causes her to do this? 'Wants to take stock' may be misleading. As with the simplest case above, and bearing in mind that we don't have an agent-cause, I think it's plausible that it's the accumulation of enough relevant considerations that causes her to discontinue the Taoist effort. At this point she begins to deterministically evaluate and weigh her new considerations into her deliberation.

There's one more possibility, which is that the deterministic phase and the 'making of the Taoist effort' phase occur simultaneously post-initiation-of-the-Taoist-effort phase. If this is the case then during the Taoist effort not only do CSCs come to mind, but what we might call DSCs (deterministically-selected considerations) come to mind also. But if we add to this another strand of thought, which would have to be not fully conscious, whereby she is not making the effort (so that DSCs can come to mind) we'd be assuming a bit too much of the agent. She'd have to be actually making the Taoist effort and not making it at the same time. And even if this is possible, the compartmentalization required would allow us to consider these processes in isolation from one another, and assess the agent's control relevant to them separately.

¹³ Actually, this is not so clear. Kane (1999, 117) reminds us that he postulates efforts to keep reasons before one's mind in the face of temptations to self-deception or backsliding. But I doubt the agent could be making these two indeterminate efforts simultaneously, i.e., an effort to relax her conscious control *and* an effort to hold on to a consideration. They're opposing tendencies.

3.6e Other Indeterministic Efforts

The fact that Kane spends relatively little time explaining how indeterministic efforts other than the Taoist effort might enhance an agent's control over a deliberative process perhaps indicates a belief on his part that these other efforts are of comparatively little significance in terms of the control required for deliberation. But even if this is so, our search for Ultimate Control on Kane's account must be thorough, so we should not dismiss these other indeterministic efforts too quickly. The aim of this section is to try to see what kind and degree of control these efforts might involve, and to assess this control in terms of our requirement of Ultimate Control established in sections 1 and 2. Once again, the indeterministic efforts in question are: '(1) efforts not to quit deliberating too soon or not to choose too hastily, (2) efforts to focus attention on considerations that have come to mind in order to determine all their consequences, (3) to avoid suppressing relevant considerations that one may be resistant to considering, (4) to galvanize oneself to pursue relevant sources of new information, and (5) to avoid deceiving oneself (or rationalizing about) the relevance of various considerations or their consequences' (1996, 167).

I have established in section 5b that agents would have diminished AD control and no Ultimate Control over the initiation of these efforts. So we should begin by looking at what kind of control an agent might have by way of the efforts post-initiation. Given what we know of Kane's account we can assume that in order for there to be a possibility of an agent's exercising ultimate control by virtue of making these efforts it must be the case that there is tension in her will as a result of her trying to make two incommensurable sets of reasons prevail. So, to take the effort not to quit deliberating too soon as an example, she must be both trying to have her reasons for

continuing to deliberate and her reasons for quitting prevail. This tension will give rise to indeterminism, making it the case that it is undetermined which reason will prevail, and whether she will thus go on deliberating or quit. But when one of the reasons does prevail, she will exercise ultimate control over its prevailing. This description seems fairly straightforward. But besides my difficulty in understanding just what kind of control this would be (given that she does not agent-cause the prevailing of one of the reasons, and that, due to the indeterminism, she does not antecedently control which one prevails), I am prepared to concede that the agent does exercise ultimate control at this point. The important question is whether this is significant in terms of the kind of control required for deliberation. I established in section 1 that the kind of control at issue is the ability of an agent to be the ultimate source of her thoughts, in the sense that they are not merely causal consequences of previous thoughts and other events. In section 2 I described this requirement as a belief in one's own authorship of a thought, and I gave some real and imagined examples as a way of making it plausible that agents who (for whatever reason) begin to question whether a thought has really been produced by them, stand in a detached, external relation to that thought, and cannot utilize it in a process of deliberation. With this in mind, we are in a position to see whether these other indeterministic efforts could give agents reason to believe that they have this kind of control.

The effort not to quit deliberating too soon would result in the agent's exercising ultimate control over the prevailing of whichever of her reasons (to keep deliberating or quit) indeterministically prevails, due to that event's being the end trajectory of the agent's effort of will to have it prevail. Let's assume that the reason that prevails is the reason to continue deliberating. At this point the agent's deliberation will cease to be interrupted by thoughts of quitting, at least for the

moment, and she will continue to deliberate. So would this instance of ultimate control be helpful in terms of her sense of authorship over her thoughts relevant to the deliberative process? I think the answer must be that it wouldn't, or wouldn't very much. The precise moment at which the agent exercises ultimate control is the moment at which she has overcome what might best be described as a temptation to quit deliberating. But this event is not really a part of the deliberative process itself; it has no content of its own, and doesn't seem to relate to the content of any particular deliberative process, i.e., the particular practical uncertainty the agent is trying to resolve and the particular beliefs, desires, inferences, etc. relevant to resolving it. To use Kane's own example to illustrate my point, 'Jane may be tired of deliberating about [where to go for] her vacation, but, recalling a previous disastrous vacation, she may make the effort... to go on deliberating rather than decide too hastily' (1996, 167). As a result of this she may believe that she is the author or ultimate source of her action to go on deliberating. But it is unclear how this would help her to stand in an internal relation to the thoughts relevant to working out what to do about her vacation. At best she would stand in an internal relation only to the prevailing of her effort to keep deliberating.

It is unnecessary to analyze each of the other kinds of indeterministic effort (2-5) that Kane claims could enhance the agent's control over the deliberative process, as the general point above seems to apply to each of them. Therefore, I think it is safe to ignore these other efforts.

3.7 Intention-Formation

Although I am unaware of any explicit discussion of intention-formation by Kane in his account of deliberation, this is perhaps the part of a deliberative process

where control of some kind, and certainly Ultimate Control, would be most desirable. That is because the formation of an intention seems to be the final point in deliberation, where there is no further practical uncertainty on the part of the agent, and the process comes to an end. It is perhaps relevant that Kane says that indeterminism doesn't give us more compatibilist (AD) control, but it does give us a different variety of control, i.e., ultimate control. In self-formation, he claims, we can't determine what we'll choose before we choose it, but we can determine what we choose when we choose it (1999, 115). This seems intended to be relevant, at least primarily, to moral and prudential decisions, as here the indeterminism plays a more central role in the decision, i.e., actually makes the difference as to what it will be. But can it also be applied to intention-formation, i.e., if this is an 'effort'?¹⁴

It seems unlikely that intention-formation is directly intended or willed (Brian O'Shaughnessy (1980, 298, 301) makes this point, and I have endorsed it in chapter 2).¹⁵ And it's very doubtful that an agent would be simultaneously trying to form an intention and not form that same intention. So the only tension there may be would result from the agent's simultaneously trying to form two different, incompatible intentions, e.g., to go to Hawaii for her vacation and to go to Colorado for her vacation. But by the stage in a deliberative process when an intention is formed the

¹⁴ It may be thought that 'deciding' and 'forming an intention' are semantically or extensionally equivalent. I do not believe this to be the case, however. For one thing, 'decision' can refer to more than just the point at which a process of deliberation comes to an end due to the formation of an intention – it can refer to the whole process (for example, "I am in the process of making a decision"). The term 'choose,' I believe, is similarly equivocal. But even if 'decision' were to refer only to the end point of a deliberative process, O'Shaughnessy seems correct to claim that not all intendings involve a decision: "Thus, I see a cup begin to topple off the table and lurch forward wildly in an attempt to catch it" (1980, 297). Here it seems that this 'instrumental act striving' is an expression of an intention (to catch the cup), yet at no point was I uncertain of what to do. Therefore, 'no decision was reached in forming the intention that found expression in my act.' Clarke (2003, 3) makes a similar claim.

¹⁵ See Clarke (2003, 3) for an opposing view. Clarke claims that we can form intentions by intending to form them, but thinks that this does not involve a regress, as the two intentions will have different content – the content of the prior intention being, for instance, to make up one's mind (26).

agent would favor reasons for the alternative on which the intention is based,¹⁶ and combined with the overriding ensuring/avoiding motivation which (I argued in chapter 2) motivates the entire process, an intention would more than likely just occur. Exactly how intentions occur as results of deliberative processes is unclear, but perhaps they are caused to form by a combination of the overriding ensuring/avoiding motivation and the relative strength of the reasons for one of the alternatives. Nevertheless, without some additional type of factor by virtue of which the agent could exercise Ultimate Control over this part of the process (excluding agent-causation), we have no reason to think that she would have anything more than AD control over it. And even AD control would be unlikely given that it requires an agent to be able to guarantee or determine an outcome *before* it occurs, which would make little sense in the context of an agent trying to work out what to do.¹⁷ On the other hand, if none of the agent's

¹⁶ Absent time constraints, it would seem irrational or at least capricious for an agent to form an intention during an earlier phase of the deliberative process, for example, the Taoist effort, as the agent has not yet weighed her alternatives and yet is in a situation where this is clearly required. And for this reason it seems unlikely that the agent would have any sense of control over its formation. (This would be similar to her forming an intention as a result of vacillating between alternatives. See below.) Even if she did have a sense of control over it, it is implausible that she could have ultimate control over it, as this would require her to be, for example, directly trying to form an intention to go to Hawaii and directly trying to form an intention to go to Colorado (in order for the indeterminism relevant to this decision to be present), *as well as* trying to be receptive to new considerations that will help her make the decision about whether to go to Hawaii or Colorado. Her trying to form an intention during the part of the process where she is weighing her reasons for each alternative (which is not explicitly discussed by Kane) would result in a similar problem, i.e., it is implausible that a person could, or would even try to undertake these parts of a deliberative process simultaneously.

¹⁷ We must not ignore the possibility that an agent might have resolved her practical uncertainty, i.e., worked out what to do, but not yet formed an intention to do it. In this case it is not inconceivable that the agent has AD control over the formation of the intention, as she would be able to determine what the intention would be before it is formed. Clarke (2003) holds a view like this. He claims that 'making an evaluative judgment is not yet forming an intention; it is not yet committing oneself to perform the action that one thereby judges best' (114-15). I do not think this view is correct. I think that if an agent has not formed an intention to do *A*, then the agent has not really decided that doing *A* is best; at least she has not made an all-things-considered judgment that it is best. But I cannot argue for this here. Further, it is not clear that the expressions 'having worked out what to do' and 'having judged an action to be best' are equivalent, so my disagreement with Clarke may be superficial. But even if Clarke is correct, and it is possible for an agent to have AD control over the formation of an intention in a case where she has already worked out what to do, or worked out what is best to do, she could not have ultimate control over the formation of the intention. If she has already worked out what to do there would no longer be any tension in her will giving rise to the indeterminism required for ultimate control.

reasons are stronger, and yet she is directly trying to form an intention rather than continuing to adduce and weigh reasons, she would be vacillating, which is not deliberation and couldn't really enhance her control over the formation of the intention (I discuss vacillation briefly in chapter 2). In either case, if an intention just occurs there would be no veridical sense of control on the part of the agent. Further, if an agent believes that this is how she will form an intention, then dissociation from the process is likely to result.

3.8 Simple Breakdown

We now need to remove all aspects of compatibilist-level (AD) control from Kane's account of deliberation in order to find out what kind of control might remain. It looks like the only place ultimate control could be relevant in any important way is during the Taoist effort. My analysis shows that here the agent exercises ultimate control over her relaxing her mind so that CSCs can indeterministically occur to her. But this is perhaps not entirely accurate. The analysis actually showed that CSCs occur only during the stage where the agent is in tension, i.e., when she is *trying* both to relax and not relax. This tension is required for the indeterminism essential for the production of CSCs. So ultimate control cannot be exercised at the point when the CSCs occur, but only when one of these efforts prevails. The agent exercises ultimate control over which one prevails *when* it prevails. Therefore, she can never exercise ultimate control during the part of the process when CSCs occur. The precise point at which ultimate control would apply is the point at which the agent has succeeded (or failed) in relaxing her mind, and is no longer trying to do anything, except maybe the ensuring/avoiding I claim would motivate the whole process. In other words, double trying is resolved into the success of one or the other of the tryings. Now the

deliberator is going to be in this relaxed 'ultimate control state' very momentarily before she returns to deliberating in the familiar way (adducing and weighing reasons deterministically). At this point, the process would somehow start up again or loop back to the beginning.

To make this clearer, we can break the process down into its key phases as follows.

1. Deterministic adducing and weighing of reasons. The agent has AD control, but no ultimate control.
2. Taoist Effort to relax is initiated. The agent has no ultimate control over the initiation, and diminished AD control over it.
3. The agent is making an effort to relax and to not relax, which is the tension that produces the CSCs. The agent has no AD control over the occurrence of CSCs. She has some AD control over making the effort. She has no ultimate control, as this is only relevant in the post-tension stage when one of her competing reasons for relaxing and not relaxing prevails.
4. The agent succeeds at relaxing (relinquishing AD control). There is no longer tension in her will, so there are no CSCs occurring. She has ultimate control, but only over her succeeding at relaxing AD control. If deliberation is to continue, the agent would now presumably return to phase 1 (at least the weighing part). It must also be possible, as success/failure is indeterministic, that she fails, i.e., her reason for not relaxing prevails. Here she would have ultimate control over failing to relax. Unless she now either forms an intention and thereby resolves her practical uncertainty, or ceases to deliberate due to some other factor, she would return either to phase 1 in

order to either continue to weigh the reasons she has, or to phase 2 in order to make another Taoist effort to find new ones.

What's unclear is how the agent moves from phase 4 back to phase 1 or 2. I've already determined in section 6c that the effort involved in phase 3, i.e., the agent's trying both to relax her mind and to not relax it, would continue as the result of the effort (the CSCs) occurs. But if the effort in phase 3 is continuous, then phase 4 would never occur, i.e., the agent would never succeed at either relaxing or not relaxing her mind, as by Kane's account ultimate control is only exercised at the precise point at which one of the agent's reasons prevails, and there would be no further tension at this point – the tension comes about from the agent's trying to resist familiar ways of thinking. Kane does say that the making of the effort is a complex neural process involving circulating impulses in a recurrent neural network. So whatever precisely this means, it looks as though the agent might move through phases 1-4 very quickly. But they can't occur simultaneously, as they're not compatible. No matter what the nature of the complexity, we can't have two incompatible activities going on in the one neural network. So it actually looks like the effort *cannot* be made continuously, but only repeatedly, and very quickly.

So my conclusion, that we can only ever exercise ultimate control over our succeeding or failing at our effort to relax (relinquish AD control), seems correct. And our ultimate control over the success (or failure) is, in a sense, a mere byproduct of this dual effort. We can have ultimate control only at stage 4, and only over our succeeding at either relaxing AD control or not relaxing it. It does not seem that this can be of much help in securing more control over a process of deliberation than an agent would have if determinism were true.

One final point about my overall argument requires clarification. In section 2 I claimed that my argument there did not conclusively establish that a deliberating agent requires a sense of authorship (ultimate control over the initiation) of her thoughts relevant to the deliberative process in addition to a sense of ownership of them. It should now be clear that even this restricted condition cannot be met by Kane's account. The explanation of this is as follows. If it is correct that believing that one is causally determined is incompatible with believing that one has Ultimate Control over one's thoughts, then we are not warranted in simply assuming that the falsity of determinism would be sufficient for this kind of control; we need an indeterministic account that provides reason to think that we have it. Now, even if the sense of control required for deliberation is control in the ownership sense alone and not also the authorship sense, and thus our search for control in an indeterministic account of deliberation is restricted to this kind of control, Kane's account does not provide it. Leaving aside the kind of control that is available if determinism is true (AD control), we have seen that the only kind of control a deliberator can exercise on Kane's account is what he refers to as 'ultimate control.' Although I have not been able to ascertain precisely what this kind of control amounts to, my analysis of Kane's account showed that it could only be exercised at a point in the deliberative process where it could not plausibly be of any help to a deliberator who needs to believe that she exercises control of her thoughts even in the ownership sense alone. She has ultimate control only over her succeeding or failing to relax her mind, and at this point in the process there are no new ideas relevant to her practical uncertainty occurring to her, and she is not weighing her reasons for action or forming an intention as to what to do.

3.9 Conclusion

It seems that all Kane's account can add to a deterministic conception of deliberation is the diminution of control. I have tried to establish in sections 3.2 and 3.3 that an agent cannot deliberate unless she believes herself to have Ultimate Control over the initiation of her thoughts relevant to the deliberative process. We have seen that Kane's account cannot provide the basis for an agent's belief that she possesses this kind of control. With the background of chapter 2 (i.e., that a deterministic conception won't plausibly allow for deliberation) it should be fairly clear that if an agent attends to her deliberative situation closely, then whether or not she believes herself to be causally determined, it will be unlikely that she will be able to consider herself to have Ultimate Control over her thoughts. Therefore, it appears that once the question of control arises, the only way an attentive agent can deliberate is with the belief that she is an agent-cause.

As mentioned above, I have used Kane's account due to its detail and comprehensiveness with respect to the process of deliberation, but I have not shown that other non-agent-causal libertarian accounts¹⁸ could not fare any better at securing the control required for deliberation. In particular I have not examined non-causal accounts.¹⁹ Although I am unable to go into any detail concerning these accounts, a

¹⁸ For example, Mele (1995) and Ekstrom (2001). These are examples of 'deliberative-indeterminist' event-causal accounts, where the indeterminism is posited at a stage in the deliberative process prior to intention-formation. According to Mele, we can exercise 'ultimate control' over the making of a decision only if beliefs relevant to that decision come to mind indeterministically during the deliberative process. On Ekstrom's account, an agent can act freely only if the preference causing her intention to act is generated by considerations that occur to her indeterministically. Kane's account of deliberation can also be seen as a deliberative-indeterminist account due to the occurrence of CSCs prior to decision and intention. Other event-causal theorists propose that a free action or decision must be indeterministically caused by its immediate causal antecedents, for instance, the agent's reasons for acting. Kane's (1996) account of moral and prudential decisions is the most well known of such accounts.

¹⁹ For example, Ginet (1990, 1997, and 2007) and Goetz (2008). Noncausal accounts require that there be some undetermined element in the causal production of an action in order for the action to be free. For some noncausalists this amounts to the requirement that the action not be caused at all, and for

general point arising from the foregoing discussion may be illuminating. It seems that the main difficulty faced by non-agent-causal libertarian accounts with respect to the control required for deliberation relates to a more general and commonly recognized difficulty for such accounts. In the absence of an agent-causal relation (a factor which gives rise to problems of its own) it's difficult to see how the indeterminism in such accounts could serve to enhance an agent's control over her thoughts and actions from what it would be if such accounts were purely deterministic. Although libertarian accounts seem to have an advantage over purely deterministic accounts in that they provide for a range of possibilities of thought and action ('alternative possibilities'), the difficulty lies in understanding how agents are able to control which of the possible thoughts and actions they produce, in a way that allows them to exercise greater control over them than would be available on a deterministic account. Although proponents of these accounts have provided detailed responses to this criticism, what is needed in terms of deliberation from a psychological standpoint is a coherent notion of agent-control that is distinct from the AD control available on deterministic accounts, and that could enhance an agent's sense of her control over her thoughts relevant to the deliberative process. Kane's ultimate control can be seen as a

others that the action is not deterministically caused. All noncausalists seem to agree that free actions do not need to have any internal causal structure. According to Ginet, every action either is or begins with a causally simple mental action, i.e., an action that 'does not contain within itself two distinct, causally related events' (1990, 12). This differs from an unbidden mental occurrence due to its '*actish* phenomenal quality' (13), which is explained as follows. '[T]he simple mental event of my volition to exert force with a part of my body phenomenally seems to me to be intrinsically an event that does not just happen to me, that does not occur unbidden, but it is, rather, as if I make it occur, as if I determine that it will happen just when and as it does.' This phenomenal quality is sufficient for the mental event's being an action. The only further condition for the action to be a *free* action of mine, is that I am its subject, and I am not in turn controlled or determined by something else, or causally necessitated by antecedent states and events (1997, 89). Goetz denies that a choice, or any mental act, has an intrinsic actish feel. He claims that the active nature of a mental event consists solely in its being the exercise of a mental power (2008, 11). His view also has a distinctive teleological aspect: A choice is 'an intrinsically active and essentially uncaused mental action that is ultimately and irreducibly explained teleologically by a reason.' Such a choice is free only if it's the case that when making it for a reason, R1, the agent also has a reason, R2, to choose otherwise, and the alternative choice is causally open to him (35).

potential way of providing this. But as I have tried to show, it does not successfully do so.

REFERENCES

- Ayer, A. J. (1954). *Philosophical Essays*, London: Macmillan.
- Campbell, John (1999). 'Schizophrenia, the Space of Reasons and Thinking as a Motor Process', *The Monist* 82, 609-625.
- Chisholm, Roderick (1964). 'Human Freedom and the Self', in Gary Watson (ed.), *Free Will*, New York: Oxford University Press.
- Clarke, Randolph (2003). *Libertarian Accounts of Free Will*, New York: Oxford University Press.
- Ekstrom, Laura Waddell (2003). 'Free Will, Chance, and Mystery', *Philosophical Studies* 113, 153-80.
- Ginet, Carl (1990). *On Action*, New York: Cambridge University Press.
- Ginet, Carl (1997). 'Freedom, Responsibility, and Agency', *The Journal of Ethics* 1, 85-98.
- Ginet, Carl (2007). 'An Action Can Be Both Uncaused and Up to the Agent', in Christoph Lumer and Sandro Nannini (eds), *Intentionality, Deliberation and Autonomy*, Burlington: Ashgate.
- Goetz, Stewart (2008). *Freedom, Teleology, and Evil*, New York: T&T Clark.
- Hoerl, Christoph (2001). 'On Thought Insertion', *Philosophy, Psychiatry, and Psychology* 8, June/September, 189-200.
- Hume, David (1748/1999). *An Enquiry Concerning Human Understanding*, Tom L. Beauchamp (ed.), New York: Oxford University Press.
- Kane, Robert (1995). 'Two Kinds of Incompatibilism', in Timothy O'Connor (ed.), *Agents, Causes, and Events*, New York: Oxford University Press.

- Kane, Robert (1996). *The Significance of Free Will*, New York: Oxford University Press.
- Kane, Robert (1999). 'On Free Will, Responsibility and Indeterminism: Responses to Clarke, Haji, and Mele', *Philosophical Explorations* 2, 105-121.
- Kapitan, Tomis (1986). 'Deliberation and the Presumption of Open Alternatives', *The Philosophical Quarterly* 36, 230-51.
- Long, A. A., and Sedley, D. N. (1987). *The Hellenistic Philosophers* v. 1, New York: Cambridge University Press.
- Mele, Alfred (1995). *Autonomous Agents*, New York: Oxford University Press.
- Nelkin, Dana (2004). 'Deliberative Alternatives,' *Philosophical Topics* 32, 215-40.
- O'Shaughnessy, Brian (1980). *The Will: A Dual Aspect Theory* v. 2, Cambridge: Cambridge University Press.
- Parnas, Josef, and Sass, Louis Arnorsson (2001). 'Self, Solipsism, and Schizophrenic Delusions', *Philosophy, Psychiatry, and Psychology* 8, June/September, 101-20.
- Pereboom, Derk (1994). 'Stoic Psychotherapy in Descartes and Spinoza', *Faith and Philosophy* 11 no. 4, 593.
- Pereboom, Derk (2008). 'A Compatibilist Account of the Epistemic Conditions on Rational Deliberation,' *The Journal of Ethics* 12, 287-306.
- van Inwagen, Peter (1983). *An Essay on Free Will*, New York: Oxford University Press.