Ishtiyaque Haji

Library of Ethics and Applied Philosophy 21

Freedom and Value

Freedom's Influence on Welfare and Worldly Value

Springer

FREEDOM AND VALUE

LIBRARY OF ETHICS AND APPLIED PHILOSOPHY

VOLUME 21

http://www.springer.com/series/6230

FREEDOM AND VALUE

Freedom's Influence on Welfare and Worldly Value

by

ISHTIYAQUE HAJI
University of Calgary, AB, Canada

Springer

Ishtiyaque Haji
University of Calgary
Dept. Philosophy
2500 University Drive NW.,
Calgary AB T2N 1N4
Canada
ihaji@ucalgary.ca

ISBN: 978-1-4020-9076-9 e-ISBN: 978-1-4020-9077-6

Library of Congress Control Number: 2008934451

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Acknowledgments

It is my great pleasure to acknowledge the help that I have received from others in the writing of this book.

I completed the last draft of the manuscript during my tenure of a 2008–2011 Social Sciences and Humanities Research Council of Canada (SSHRC) Grant. I am most grateful to this granting agency for its support.

A number of anonymous referees, both for Springer and for certain journals, provided highly instructive criticism and helpful advice regarding sections of the manuscript. I am very thankful for this help.

Parts of the following previously published articles of mine appear either verbatim or modified in the book. With permission of Routledge, Taylor & Francis Group: Ishtiyaque Haji and Stefaan E. Cuypers, "Moral Responsibility and the Problem of Manipulation Reconsidered," *International Journal of Philosophical Studies* 12 (2004), pp. 439–464; and Haji and Cuypers, "Magical Agents, Global Induction, and the Internalism/Externalism Debate," *Australasian Journal of Philosophy* 85 (2007), pp. 343–371; see also, these Journals' web site: www.informaworld.com. With permission of Blackwell Synergy: "Incompatibilism's Threat to Worldly Value: Source Incompatibilism, Desert, and Pleasure," forthcoming in *Philosophy and Phenomenological Research*. With permission of Springer: "Authentic Springs of Action and Obligation," forthcoming in *The Journal of Ethics*. With permission of Cambridge Scholars Publishing: "Obligation and Luck," forthcoming in *Essays on Free Will and Moral Responsibility*. Nick Trakakis, ed. With permission of the University of Arkansas Press: "Freedom, Hedonism, and the Intrinsic Value of Lives," *Philosophical Topics* 32: 131–151. My paper, "On Frankfurt-Type Examples," from which I have drawn first appeared in the *APA Newsletter on Philosophy and Law*, Vol. 07, No. 2, Spring 2008. I am grateful to the editors and publishers for their permission to use materials from these essays.

Many thanks to Marion Wagenaar, Editorial Assistant at Springer for her diligent work, and to Fritz Schmuhl, Publishing Editor at Springer; their valuable contributions are very much appreciated.

During the time that I have been thinking about the topics of this book, I have benefited enormously from exchanges with a number of people. I take immense pleasure in acknowledging my indebtedness to them here. They are Chris Heathwood, Michael McKenna, David McNaughton, Alfred Mele, Derk Pereboom,

and Piers Rawlings. I am especially grateful to Stefaan Cuypers who read a complete version of the manuscript.

Finally, there are two people who deserve very special mention. Michael J. Zimmerman has been a tremendous source of inspiration to me right from the beginning of my academic career. His prodigious intelligence shines through all his work. His books and papers exhibit uncompromising intellectual integrity, acute attention to detail, and amazing insight. He has stopped me from making a number of errors and suggested many, many improvements. He has given me steadfast encouragement during trying times, and has been ever ready to help. His support has been invaluable. He has been, and continues to be, my mentor and dear friend. I shall never be able to discharge the debt I owe to him.

If it had not been for Fred Feldman, I would not have gone on in philosophy. "Find a friend who can spell," he said to me sternly but with a twinkle in his eyes on returning the first paper that I drafted for him. He taught me how to write well, formulate arguments, and horn my analytic skills. He cultivated my love for philosophy. He has enriched my academic life beyond expression. Anyone who has studied Fred's numerous contributions to philosophy will recognize his genius. His works are matchless in their display of clarity, rigor, insight, and overall scintillating brilliance. If I could only think and write like Fred, I could give back a scintilla of what I owe him. Fred is very dear to me indeed.

It is to Michael and Fred that I dedicate this book with profound gratitude and respect.

Contents

Chapter 1
Introduction: On Welfare and Worldly Value

1.1 The Focus of Inquiry: The Freedom of Axiological Judgments

Hypological, deontic, and *axiological* judgments are among the many different sorts of moral judgment, each sort expressing a moral appraisal of some variety. Hypological judgments are judgments concerned with moral responsibility.[1] It is widely accepted that freedom is relevant to the truth of such judgments. We believe that a person cannot be morally blameworthy or praiseworthy for an action unless she exercises "freedom-relevant control" in its performance; in more mundane parlance, she cannot be responsible unless she acts with "free will." It has been less commonly acknowledged that freedom also bears on the truth of morally deontic judgments that have to do with moral right, wrong, and obligation. For instance, it has been argued that if an action is morally obligatory for an agent, then she could have refrained from performing it (Zimmerman 1997, Haji 2002). Relatively little thought, however, has been invested into whether the truth of *axiological judgments*—judgments regarding good or evil—presupposes freedom. I suspect that this important matter has not garnered the attention that it deserves owing to what is perhaps a prevalent assumption that freedom leaves axiological appraisals, by and large, unaffected. The central aim of this book is to dispute this assumption by arguing for the relevance of freedom to axiology.

Many distinctions within the category of axiological judgments can be made. Within the broad class of such judgments, the ones of interest in this work are, first, axiological judgments that are judgments of intrinsic value. Roughly, something is intrinsically good if it is good "in itself"; it is good "for its own sake" (the notion of being intrinsically bad is to be similarly understood). This sort of value is often contrasted with instrumental value, which is one type of extrinsic value: loosely, something is instrumentally good if it is good as a means; it contributes either directly or indirectly to the existence of something that is intrinsically good.[2]

---

[1] Michael Zimmerman (2006, p. 585, n. 1) explains that the term "hypological" is drawn from the Greek *ύπολογος*, meaning "held accountable or liable."

[2] The term "instrumental value" can also be used to refer to a type of intrinsic value. See Ronnow-Rasmussen (2002b). I make no such use of it in this work.

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

Second, the judgments commanding interest may be regarded as a subset of the class of judgments of intrinsic value. These are axiological judgments that pertain both to personal well-being (or, alternatively, individual welfare) and to the intrinsic value of worlds. Regarding the former, philosophers have long inquired into what makes a life good in itself for the one who lives it. What makes a life high in individual welfare? Many different answers have been given to this question. Hedonism, for example, is one of them. The hedonist appeals to the view that pleasure is the *good* to account for the amount of welfare that an individual enjoys. In the hedonist's estimation, the *good life* is the pleasant life. In contrast, preferentists maintain that what makes a life intrinsically good for a person is that desires of some sort are satisfied rather than frustrated within that life. Regarding the assessment of worlds, the concern is not with individual well-being but, rather, with the intrinsic value of *entire* worlds: when is one possible world intrinsically better than another?

If we suspect that freedom in some manner or other affects the intrinsic value of lives or worlds, how precisely are we to establish that this is so? To make headway, I harness the following two strategies. First, I agree with a number of theorists that every axiology—roughly, every theory of intrinsic value—specifies some items that have their intrinsic values in the most primary way (see, for instance, Harman 1967, Chisholsm 1986, Feldman 2000, Zimmerman 2001). The *basic intrinsic value states* of each axiology are the items that the axiology takes to be the most fundamental bearers of intrinsic value (Feldman 2004, p. 173). Each of these items has its intrinsic value in a nonderivative way. Think of each such item as an "atom" of value. The intrinsic value of a complex thing, such as a life, a world, or the total consequence of an action, is the sum of the value of its atoms. To appreciate what these atoms may be on hedonistic axiologies, drawing a distinction between attitudinal pleasures and displeasures, on the one hand, and sensory pleasures and pains, on the other, is helpful. A person experiences sensory pleasures at a time when she feels pleasurable sensations. Attitudinal pleasures are not feelings or sensory pleasures; they need not have any "feel." Such pleasures are always directed onto objects; they are propositional attitudes. A person takes attitudinal pleasure in some state of affairs "if he enjoys it, is pleased about it, is glad that it is happening, is delighted by it" (Feldman 2004, p. 56).[3] To take intrinsic pleasure in a fact is to take pleasure in it for its own sake (corresponding things are true about displeasures). One sort of hedonist—the sensory hedonist—proposes that the atoms of value that contribute to welfare value are episodes of sensory pleasure and sensory pain; a different sort—the attitudinal hedonist—takes these atoms to be episodes of intrinsic attitudinal pleasures and intrinsic displeasures. A preferentist, in contrast, may

[3] For more on the distinction between sensory and nonsensory pleasures, see, for example, Brentano (1969, pp. 154–155); Chisholsm (1986, p. 26); and Lemos (1994, pp. 67–73). Zimmerman (2001, pp. 195–198) proposes that attitudinal pleasures and displeasures do have an affective aspect, so an adequate account of the nature of attitudinal pleasure and displeasure must make reference to their affective aspect; and that an adequate account of the value of these attitudes must also make reference to this aspect.

take episodes of desire satisfactions and desire frustrations of one's actual desires to be the atoms. On the first strategy to uncover the freedom presuppositions of life- or world-ranking axiological judgments, we inquire into whether free "positive" atoms—free intrinsic attitudinal pleasures, for example—are intrinsically better than (or maybe not as good as) otherwise similar atoms that are unfree, and whether free "negative" atoms—free intrinsic displeasures, for instance—are not as bad as (or perhaps worse than) otherwise similar unfree atoms, the freedom at issue being the freedom that moral responsibility demands. Are there indeed reasons to believe that free intrinsic attitudinal pleasures, for instance, are more valuable than otherwise similar unfree pleasures? And if they are more valuable, are they more valuable for the world or for the well-being of the subject who receives them or for both?

I believe that this first strategy can be effectively implemented only in connection with substantive views about the *good life* or about the intrinsic value of worlds. For the most part, the substantive theories to which discussion in this work is confined are various forms of attitudinal hedonism. I shall, though, also address some varieties of preferentism and a prominent version of a "whole-life satisfaction" account of welfare that explains personal well-being by appealing to a certain "whole-life" view of happiness (Sumner 1996). In this way, the inquiry into the freedom presuppositions of the pertinent axiological judgments will be restricted but rendered more manageable at the same time.

Second, having made the prima facie case that free positive atoms (on some of the substantive views at issue) are better either for lives or worlds than otherwise similar atoms that are not free and that free negative atoms are not as bad, again either for lives or worlds, as otherwise similar unfree atoms, we may look more closely at the very constituents of the atoms. We may ask whether these constituents themselves have freedom presuppositions. As an illustration, many have supposed that deserved intrinsic pleasures are better than corresponding pleasures that are not deserved and that deserved intrinsic displeasures are not as bad as otherwise similar displeasures that are not deserved. Suppose that on some credible version of attitudinal hedonism, desert of intrinsic pleasure (or of intrinsic displeasure) is an element of an atom—a "world atom"—that pertains to the intrinsic value of worlds. An action-based pleasure (or displeasure) is a pleasure that, if deserved, is deserved on the basis of performing some action. If it can be argued that such a pleasure (or displeasure) is deserved only if it is free, and if it can further be secured that a pleasure is free only if it causally derives from a decision that is free, then we can show that freedom has an indirect impact on the intrinsic value of worlds by way of having an impact on desert and mental action.

A predominant part of my interest in uncovering the freedom presuppositions of various axiological judgments resides in the following. Determinism is the thesis that there is at any instant exactly one physically possible future (van Inwagen 1983, p. 3). On this picture, the future is not a "garden of forking paths" but is a branchless extension of the past. One of the central questions in the free-will debate is whether the freedom that moral responsibility requires is compatible with determinism. What this freedom amounts to is, of course, something that is hotly debated.

Some theorists have argued that a person has the right sort of freedom or control only if she had genuine alternatives; she "could have done otherwise" (see e.g., van Inwagen 1983, Ginet 1996). Others have suggested that a person has the required control only if she is appropriately sensitive to reasons; she would, under specified conditions, have performed some other action were apt reasons present (see, e.g., Fischer and Ravizza 1998, Haji 1998); still others have maintained that a person has the pertinent control just in case she identifies with the action's motivating desires (Frankfurt 1971); and yet others have defended the view that the germane control consists in the action being produced nondeviantly by causal antecedents such as desires, beliefs, values, and so forth that satisfy certain constraints (Mele 1995). If determinism is true (i.e., if it is true that all events are deterministically caused or, as I shall say, are "causally determined"), the "genuine" (nonrelational) facts of the past, together with the laws of nature, entail all truths. Compatibilists have argued that determinism poses no threat to judgments of moral praiseworthiness or moral blameworthiness. More generally, they defend the view that free action and moral responsibility are compatible with determinism. Incompatibilism is the denial of compatibilism. Incompatibilists, for one or more reasons, have championed the view that determinism undermines the truth of hypological judgments. Some incompatibilists have attempted to show that determinism rules out alternative possibilities. They have gone on to claim that freedom to do otherwise is necessary for responsibility, and thus have arrived at the conclusion that determinism and responsibility are not friendly partners. Other incompatibilists have proposed that if our actions originate in sources over which we have no control, then we are not the ultimate originators of these actions. Such "source incompatibilists" have suggested that "ultimate origination" is a requirement of free action and responsibility, and thus have, via this route, concluded once again that determinism and responsibility are incompatible. Libertarians are incompatibilists who believe that at least some of us, at times, perform free actions for which we are morally responsible; their view implies that if we are morally responsible, then determinism is false.

Assume, for the sake of explaining a pivotal incentive to inquire into the freedom presuppositions of axiological appraisals having to do with welfare and worlds, that incompatibilists in the free-will battle have won the day. If determinism undermines the freedom that moral responsibility requires and if it is also true that free atoms (on the substantive axiologies in question) are more intrinsically valuable than otherwise similar unfree atoms, then determinism has a nontrivial effect on intrinsic value. Further, suppose it can be shown that the freedom of the atoms on a given axiology traces to, or is derivative from, the freedom of various actions, such as (the making of) decisions. Then this opens up another venue for the incompatibilist to marshal support for the view that determinism is more menacing than it has hitherto been appreciated to be: it has a definite impact on various, significant axiological appraisals. So just as one may wonder whether determinism is compatible with the truth of hypological judgments, so, too, one may wonder whether determinism is compatible with the truth of various axiological judgments.

1.2 Synopsis

There is a trio of major themes that this book explores or develops. I outline these themes first. Then I follow with a chapter-by-chapter summary.

1.2.1 Major Themes

The first theme focuses on freedom and personal welfare. I advance a prima facie case for the view that just as freedom has an impact on the truth of judgments of moral responsibility, so freedom has an impact on the truth of judgments of welfare. While positive atoms (e.g., episodes of pleasure on hedonistic views or desire satisfactions on preferentist views) enhance intrinsic value, negative atoms (episodes of displeasure on various hedonistic views or desire frustrations on preferentist views, for instance) mitigate such value. I argue that on promising hedonistic views or preferentist views or whole-life satisfaction views of welfare, free positive "life atoms"—atoms that contribute to welfare value—are better than otherwise similar unfree atoms, and free negative life atoms are not as bad as otherwise similar unfree atoms. For example, *freely* taking pleasure in something is better than an otherwise similar unfree atom, and freely taking displeasure in something is not as bad as an otherwise similar unfree atom. So a life with free atoms of value, other things being equal, is better in itself for the person than an otherwise similar life containing unfree atoms. If this is all true, then determinism (given incompatibilist assumptions) should have a negative impact on personal welfare.

The second major theme concentrates on freedom and worldly value. I make a case for the view that freedom has a marked impact on the truth of judgments of worldly value. I argue that free positive world atoms—atoms that enhance worldly value—are better than otherwise similar atoms that are unfree, and that free negative world atoms—atoms that assuage worldly value—are not as bad as otherwise similar atoms that are unfree. So a world with free atoms of value, other things being equal, is intrinsically better than an otherwise similar world with unfree atoms. Again determinism (on incompatibilist assumptions) should have a negative impact on worldly value.

A prevalent position in the recent literature on freedom and responsibility is that lack of free will is not such a bad thing (Waller 1990, Honderich 1993, 2002; Pereboom 2001, 2002; Double 2004). The third primary theme makes a case for the view that living without free will does, contrary to this position, have serious hitherto unnoticed costs that are associated with the intrinsic value of lives and worlds.

One central project in moral philosophy is the attempt to identify the good life—the attempt to identify the features in virtue of which a person's life is good in itself for that person. Many theories about rationality also presuppose the notion of a good life. On these views, the concept of rationality is to be explained by appeal to the concept of individual well-being. What's rational for you to do is what will most enhance your well-being (usually weighted for probability). The first major theme—what influence does free will (or its lack) have on personal well-being?—is directly

relevant to these fundamental issues. Another central project in moral philosophy is the attempt to explain the concept of obligation in terms of the intrinsic value of worlds. The second major theme—what influence does free will (or its lack) have on worldly value?—bears significantly on this project. A project pivotal in the literature on free will is whether being deprived of freedom is such a bad thing. The third major theme—what are the implications of the view that, other things being equal, freedom enhances welfare value and worldly value for various issues in the free-will debate?—speaks to this project.

1.2.2 Chapter-by-Chapter Summary

The book is organized as follows. Refer to an axiology that ranks lives (as I shall say)—an axiology that responds to what makes a person's life good in itself for that person—as a "life-ranking axiology" and to an axiology that ranks worlds as a "world-ranking axiology." I begin in Chapter 2 with a clear formulation of a relatively simple version of intrinsic attitudinal hedonism that derives from Fred Feldman's insightful work on hedonism. According to the simple theory, which, in one incarnation at least, is a life-ranking axiology, the atoms of value that it selects (the life atoms) are episodes of intrinsic attitudinal pleasures and intrinsic displeasures. I summarize pivotal reasons to prefer this theory to more traditional forms of hedonism on which the life atoms are episodes of sensory pleasure and pain. Departures from the simple theory are introduced to handle various objections to the theory. Some of the modified versions resulting from the simple theory, for example, adjust the values of episodes of intrinsic pleasure (and intrinsic displeasure) in accordance with whether or not the objects of one's pleasures are true, or in accordance with whether or not the objects of one's pleasures deserve to have pleasure taken in them.

In Chapter 3, conscripting the first of the strategies that I outlined above, I advance considerations in support of the view that a life-ranking axiology, such as the simple theory, should adjust the value of its life atoms to reflect the extent to which they are free. I explain that if a pleasure is free, it is only indirectly so: its freedom traces to, or derives from, the freedom of decisions. I construct various freedom-sensitive accounts of intrinsic attitudinal hedonism—life-ranking accounts that adjust the values of intrinsic pleasures (and intrinsic displeasures) in accordance with whether or not the pleasures (and displeasures) are free. I then defend selected versions of these freedom-sensitive views against various objections. I settle on what I deem is the best contender. On this freedom-sensitive contender, free intrinsic attitudinal pleasures are better than otherwise similar unfree ones, and free intrinsic displeasures are not as bad as otherwise similar unfree ones.

Presupposing an account of attitudinal hedonism as a life-ranking axiology, freedom may have a direct bearing on welfare if it is indeed the case that free pleasures are better than otherwise similar unfree ones and free displeasures are not as bad as otherwise similar unfree ones. In Chapter 4, I explore whether freedom has a somewhat oblique impact on well-being. I pursue this line of thought: suppose de-

served intrinsic pleasures are better than otherwise similar pleasures that are not deserved, and that deserved intrinsic displeasures are not as bad as otherwise similar displeasures that are not deserved. Owing to its being plausible that freedom influences attributions of desert, it is reasonable to think that freedom influences the intrinsic values of deserved (or undeserved) pleasures and deserved (or undeserved) displeasures. Suppose, further, that the atoms pertinent to evaluating *welfare* include information about the extent to which episodes of intrinsic pleasures and intrinsic displeasures are deserved. Then, via having an influence on the desert of pleasures and displeasures, freedom affects welfare. The official position I defend is that with some version of attitudinal hedonism as the life-ranking axiology, it is plausible that the value of life atoms be adjusted to reflect the extent to which pleasures and displeasures are deserved or not deserved, *if* it is also plausible that the value of such atoms be adjusted on the basis of whether the objects of the pleasures or displeasures are true or *if* it is plausible that the value of these atoms be adjusted on the basis of whether the objects of the pleasures (or displeasures) are objects that deserve to have pleasure (or displeasure) taken in them. I confess, though, to being drawn to the view that desert, really, has an impact on the value of worlds and not on the value of lives.

To understand the influence of freedom on episodes of intrinsic attitudinal pleasure and displeasure and to apprehend freedom's influence on desert of such pleasures and displeasures, it is necessary to say something about the freedom of our decisions. This is partly because, as I have registered, if attitudinal pleasures (or displeasures) are free, they are only indirectly so: their freedom derives from the freedom of elements in our mental life with respect to which we are directly free. On the view that I favor, these elements are decisions. I propose that a decision is free if and only if its agent has control in making it and it causally arises nondeviantly from springs of action, such as desires and beliefs, that are "truly the agent's own" or "authentic." Thus, if an episode of attitudinal pleasure (or displeasure) is free, it derives from springs of action that are authentic. I devote Chapter 5 to giving an account of the authenticity of actional springs.

In Chapter 6, I highlight the impact of freedom both on welfare and on the value of worlds by inquiring into whether the truth or falsity of determinism would have a bearing on welfare or worldly value. First, I discuss several arguments to the effect that if source incompatibilism is true, determinism has a definite impact on well-being and the value of worlds, on the assumption that freedom-sensitive versions of attitudinal hedonism to be introduced are either life- or world-ranking axiologies. Then I discuss the impact, if any, that compatibilist and libertarian views have on welfare and the value of worlds, again assuming that apt life- and world-ranking freedom-sensitive versions of attitudinal hedonism are true.

The influence of specific versions of world-ranking freedom-sensitive attitudinal hedonism on moral obligation is the broad topic of Chapter 7. It has seemed very natural to many that our fundamental moral obligation is to do the best we can—to make the world as good as we can. Assume that some version of freedom-sensitive attitudinal hedonism—specifically, a version on which the values of world atoms are sensitive both to freedom and desert—provides the axiology that yields a ranking

of possible worlds in terms of their intrinsic values. Given these presuppositions, in this chapter, I first argue that determinism has a nontrivial impact on obligation *owing to its impact on the value of world atoms*. On some accounts of a freedom-sensitive axiology, determinism undermines obligation; on other accounts, although determinism does not impugn the truth of morally deontic judgments because of its impact on the value of world atoms, determinism nevertheless has an effect on obligation as a result of this impact. On route to securing these conclusions, I formulate various theories of freedom-sensitive attitudinal hedonism as world-ranking axiologies and advance reasons for preferring one of them.

Totalism is the view that the value of a world is the sum of the value of the world atoms that it contains. The preferred world-ranking axiology introduced in this chapter (Chapter 7) that adjusts the values of its atoms for both freedom and desert is a totalistic axiology. Derek Parfit has argued that any totalistic axiology succumbs to the so-called problem of the "repugnant conclusion." Briefly, the problem is this: if there are very many recipients of value in a world, a totalistic axiology rates the world as highly good even though each of its recipients receives a nominal share of the value. In addition, such an axiology rates a world with relatively fewer inhabitants, each of whom receives a far larger share of value but whose worldly value is lower than the worldly value of the first, as worse than the highly populous world; it does so even though each denizen of the less populous world fares much better (Parfit 1984, p. 383). Many find this conclusion repugnant: surely the less populous world is superior. Addressing this objection is the second major undertaking in Chapter 7. I argue that the objection does not impugn totalism.

Finally, I confirm that on the freedom- and desert-sensitive world-ranking axiology that I prefer, there is a nontrivial connection between moral obligation, on the one hand, and the inauthenticity of our springs of action, on the other: it is possible that, *in virtue of an action arising from inauthentic springs*, its agent is not morally obligated to perform this action.

Derk Pereboom has recently advanced a powerful defense of a position in the free-will debate that he calls "hard incompatibilism." According to this position, with the exception of agent-causal accounts of free action or moral responsibility, no compatibilist or libertarian account is true. Pereboom believes, further, that since there is no scientific evidence to corroborate the view that we are agent-causes, we are, in effect, without free will. A significant supplement to Pereboom's hard incompatibilism (elements of which are shared by some other incompatibilists) is that, contrary to what many have believed, a life devoid of free action would not be as detrimental as it has often been made out to be and, in certain respects, such a life would even be beneficial (Pereboom 1995, 2001, 2002). Pereboom for instance, argues that though lack of this freedom undermines appraisals of moral praiseworthiness and moral blameworthiness, its absence leaves intact *other* sorts of moral appraisal, such as appraisals of moral obligation, right, and wrong. In Chapter 8, I undertake to show that a hard incompatibilist who accepts the preferred freedom- and desert-sensitive world-ranker discussed in the previous chapter does, pace Pereboom, incur costs having to do with both the repugnant conclusion and moral obligation.

In Chapter 9, I discuss yet another cost of hard incompatibilism. Invoking hard incompatibilist assumptions, I argue, first, that determinism (and so hard incompatibilism) is incompatible with the truth of judgments of practical reason; if determinism is true, for instance, then no one can have an objective, *pro tanto* reason to do anything. I propose, next, that there is a connection between values expressed by states of affairs, such as *Simon's taking pleasure in drinking wine is intrinsically good for Simon* and *Al's taking displeasure in sitting in the sun is intrinsically bad for Al,* and practical reasons that is pertinent to whether determinism has an influence on such values: if, for instance, the former state of affairs obtains, then Simon has a (possibly defeasible) *pro tanto* reason to bring about the state of affairs *his being pleased in drinking wine*. On certain incompatibilist presuppositions, as determinism precludes the having of practical reasons, determinism precludes such values as well. I also outline an argument for the view that determinism undermines moral obligation owing to determinism's undermining the intrinsic value of worlds (again, appealing to certain incompatibilist assumptions and assumptions about the correct life- and world-ranking axiologies).

Adding to the presuppositions, first, that we morally ought to do the best we can and, second, that a freedom- and desert-sensitive axiology is the world-ranking axiology, the additional presupposition that one cannot, at a time, have a moral obligation to do something unless one has, at that time, a "genuine" alternative, in Chapter 10 I argue that moral obligation is subject to luck in a fashion in which I explain because the intrinsic value of worlds is itself subject to luck. The influence of "axiologically laden" luck on well-being is also touched upon briefly.

Some may think that freedom has an influence on intrinsic value only if the life- and world-ranking axiologies are varieties of attitudinal hedonism. In Chapter 11, I show that this claim is mistaken. Having formulated what is a fairly sophisticated version of preferentism, I argue that even this version yields implausible ratings of lives in virtue of not being sensitive to considerations of freedom. The freedom concerns an aspect of the freedom of desires: the desires are free in that they must be truly one's own or authentic.

Finally, in Chapter 12, toward further motivating the view that freedom is a constituent of any life- or world-ranking axiology, I discuss L.W. Sumner's insightful account of welfare. Sumner's account builds on a "whole life satisfaction" theory of happiness. Interestingly, Sumner attaches an authenticity requirement on happiness, though details of the requirement are not supplied. His account of welfare, thus, inherits this authenticity requirement. In roughly hewn strokes, Sumner proposes that a person's life is going well for her if and only if she judges—she makes a happiness judgment—that, on balance, her life measures up favorably against her "life ideal"; this judgment of hers is relevant, sincere, considered, informed, and autonomous, and she feels satisfied or fulfilled with her life (Sumner 1996, Chapter 6). I argue that if Sumner's axiology is the correct life-ranking axiology, then if hard incompatibilism is true, our lives are shorn of welfare value.

Chapter 2
Attitudinal Hedonism

2.1 Simple Intrinsic Attitudinal Hedonism

I begin by introducing a straightforward, relatively simple version of attitudinal hedonism. To formulate this simple theory, a few things about the basic intrinsic value states—the atoms of value—on this axiology first require elucidation. I assume in what follows that basic intrinsic value states are all finely grained propositional entities; they are states of affairs. There is no general consensus on the truth of this assumption. So, for example, Noah Lemos describes what he takes to be the bearers of intrinsic value as abstract, but it appears that he does not believe that these abstract objects are states of affairs (Lemos 1994, pp. 20–31). Moore, at times, talks of individual physical objects, such as books, as having intrinsic value ([1903] 1962, p. 3). Michael Zimmerman proposes that the bearers of value are concrete states of individuals (2001, pp. 52–64). Fortunately, this dispute in axiology concerning the ontology of the bearers of intrinsic value is of no consequence to one of the key themes of this book—exposing the freedom presuppositions of the atoms of value of both life- and world-ranking axiologies.

We now record the following about basics. First, each of the basic intrinsic value states on an axiology has its intrinsic value in a *nonderivative* way. The atoms do not, for example, have their intrinsic values because they are conjunctions of smaller items that are intrinsically good (Feldman 2004, p. 173). Further, assume that there are minimal, nonoverlapping intervals of time. Explicating, Feldman writes

> To say that time is discretized is to say that time is broken down into a sequence of tiny intervals. Each of the intervals is of very tiny duration: the collection of them is exclusive, in the sense that no two of them overlap temporally; and the collection is exhaustive, in the sense that there is no period of time that falls outside the collection. If you have all the minimal intervals, you have all the time there is (2004, p. 174).

The assumption that time is discretized reveals another way in which a state of affairs might have its value derivatively. Suppose a sensory hedonist believes that the positive atoms of intrinsic value are episodes of sensory pleasure; they are states of affairs in which some person feels sensory pleasure of some intensity for some duration. And suppose this sort of hedonist thinks that the negative atoms are episodes of sensory pain. If a person feels pleasure throughout a duration lasting 5 minutes, the

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

person, it seems, also feels pleasure at many shorter intervals during the 5-minute period. If we count each of these overlapping intervals, we will have overestimated the amount of pleasure he feels during this period. Temporal discretization goes some way toward enabling us to cope with this problem. Basics are all alike with respect to duration: each basic state is about someone's pleasure at a precise, named, minimal unit of time. A 5-minute episode of pleasure is a nonbasic state containing a large collection of such basic value states (2004, p. 174).

In addition, a generalized state of affairs such as *everyone feeling lots of pleasure at t*, where *t* indicates one of the minimal units of time is not a basic state. On a hedonistic axiology, this state is intrinsically good in virtue of its being the case that specific facts about the pleasures of specific individuals—"everyone"—are true (2004, p. 174).

Second, the basic intrinsic value states on an axiology must contain no more than the elementary bits of information that bear on intrinsic value according to the axiology in question; basics are stripped off all extraneous information. On the version of sensory hedonism to which we have just alluded, the positive atoms would all look relevantly like the following:

B1: Bob feeling sensory pleasure of intensity +7 at noon on October 16, 2001.

The essential bits of information in basics of this sort are who is feeling the sensory pleasure? How much pleasure is he feeling? When is he feeling it? The added bit of information that *while sipping a soda* Bob feels the seven hedons of pleasure that he does is superfluous to a basic's information; the intrinsic value of the basic on this axiology does not depend on this fact. Again, a basic value state contains no information beyond the information that is strictly relevant to its determination of intrinsic value.

To underscore this vital point, Feldman invites us to reflect on the logical structure of basics. Regarding B1, letting "t" stand for the indicated date, and letting "b" stand for Bob, we can introduce "P" as a three-place relational predicate designating the feeling of sensory pleasure by someone to some degree at some time. The underlying logical structure of B can be represented in this way:

$$B1^{*} : Pb, t, +7.$$

The attribution of sensory pleasure to Bob in $B1^{*}$ is a pure attribution of sensory pleasure since the other constituents of $B1^{*}$ are all picked out by mere tags—rigid designators. The state of affairs $B1^{*}$ expresses is an attribution of sensory pleasure to a person, at a designated minimal time, to a precise and designated degree (2004, p. 176).

Third, if some state of affairs is a basic intrinsic value state on some axiology, then it must have a fully determinate intrinsic value on that axiology, and it must have it intrinsically (2004, p. 175).[1] I accept the view that the intrinsic value of a

[1] For skepticism concerning the view that a basic has a fully determinate intrinsic value, see Lemos (2006).

thing is a component of its value that depends on the intrinsic features of the thing and not on its relations.

Fourth, what makes a life or a world valuable is not the basic intrinsic value states that *exist* in that life or world but the basic value states that *obtain* in it. Assuming that B1 is a basic value state of some form of sensory hedonism, if B1 does not obtain, it contributes nothing of value to Bob's life or to the world despite the fact that it itself is intrinsically good and exists at each world (Feldman 2000, p. 323, p. 344, n. 25.)

What counts as a basic will vary across axiologies. The basics on the simple version of attitudinal hedonism are all fully determinate, precise, "pared down" attributions of intrinsic *attitudinal pleasure* and *displeasure*, all relevantly like this:

B2: At noon on Tuesday, October 16, 2001, Bob takes intrinsic attitudinal pleasure of intensity +8 in the fact that Bob's beer is frosty cold.

Letting "c" stand for the state of affairs *Bob's beer being frosty cold*, "t" for the indicated date, and "b" for Bob, introduce "P" as a four-place relational predicate that designates the taking of intrinsic attitudinal pleasure by someone to some degree (where degree is a measure of strength of attitude) at some time in some object. B2's logical structure can be represented as follows:

$$B2^{*} : Pb, t, +8.$$

In addition to pure attributions of pleasure, as in B2, there are also pure attributions of displeasure (attitudinal pain) on this axiology. Here is an example:

B3: At noon on Tuesday, October 16, 2001, Bob takes intrinsic displeasure of intensity +3 in the fact that Bob's peanuts are soggy.

Prior to presenting the simple theory, let us recapitulate. First, attitudinal pleasures and displeasures are propositional attitudes. They may involve sensations or feelings, but they need not always do so. What, precisely, is the appropriate attitude? The attitude cannot be characterized too broadly as simply a *pro* attitude—an attitude of being for or in favor of something—for one can favor something without taking pleasure in it. Rather, the attitude appears to be something like enjoyment (though, again, enjoyment that is not essentially associated with feelings of enjoyment or sensory pleasures). One can take attitudinal pleasure and displeasure in just about any kind of object, and if no sensation is involved, no feeling is involved.[2] Second, to take intrinsic pleasure or intrinsic displeasure in an object is to take pleasure or displeasure in it for its own sake. Third, whenever a person takes intrinsic attitudinal pleasure in something, he takes pleasure of some degree. "Degree" here is

[2] There is an instructive discussion on what the appropriate attitude might be, and on the connection between attitudinal pleasures (displeasures) and sensory pleasures (displeasures) in Sumner (2006). On Feldman's view, a sensation is a sensory pleasure for a person at a time *t*, if and only if this person takes intrinsic attitudinal pleasure in the fact that she is experiencing that sensation at *t*.

not to be understood as intensity of any sensation. It is to be understood as strength of attitude. Fourth, when a person takes intrinsic pleasure in something, he does so for a minimal nonoverlapping period of time. Thus, there are minimal episodes of intrinsic attitudinal pleasure and minimal episodes of intrinsic displeasure. According to the simple theory, the atoms of intrinsic value—the basics—are (minimal) episodes of intrinsic attitudinal pleasure and intrinsic displeasure all relevantly like B2 and B3. The theory—"simple intrinsic attitudinal hedonism"—may be formulated in this way:

Simple Intrinsic Attitudinal Hedonism (SIAH)

i. Every episode of intrinsic attitudinal pleasure—every basic or atom such as B2—is intrinsically good; every episode of intrinsic displeasure—every basic or atom such as B3—is intrinsically bad.
ii. The intrinsic value of an episode of intrinsic attitudinal pleasure is equal to the amount of pleasure contained in that episode; the intrinsic value of an episode of intrinsic displeasure is equal to –(the amount of displeasure contained in that episode) (Feldman 2004, p. 66).

Simple intrinsic attitudinal hedonism is compatible with its being the case that the atoms that contribute to the intrinsic value of a world ("world atoms" as we have said) may differ, in significant respects, from the atoms that contribute to the intrinsic value of a person's life ("life atoms"). We will revert to this point later. Concerning lives, clauses (i) and (ii) are to be supplemented with a third:

iii. The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and displeasure contained in the life, in such a way that one life is intrinsically better than another if and only if the net amount of intrinsic attitudinal pleasure in the one is greater than the net amount of intrinsic attitudinal pleasure in the other (Feldman 2004, p. 66).[3]

Simple intrinsic attitudinal hedonism may profitably be compared with a default version of sensory hedonism. Making some assumptions about sensory pleasures and pains, each episode of sensory pleasure contains an amount of pleasure that is a product of its intensity and duration. Call a unit of pleasure a "hedon." Stronger and longer episodes of sensory pleasure contain more hedons than weaker and shorter episodes. Similar things are true about episodes of sensory pains. Call one unit of pain a "dolor." Assume that pain is "opposite" to pleasure, so we can speak of the balance of pleasure over pain. The net sensory pleasure that a person feels during a stretch of time is obtained by subtracting the number of dolors that she feels during that stretch from the number of hedons that she feels during that stretch. We said that on the default axiology, the atoms of value are episodes of sensory pleasure and sensory pain. The theory can be stated in this way:

---

[3] Sumner (1996, pp. 103–112) invites us to think of this sort of hedonism as an improvement over the default theory. In the terminology that he adopts, "suffering" denotes displeasure (or something like it) and "enjoyment" refers to attitudinal pleasure (or something akin to it). On suffering and enjoyment, also see Sumner (1996), pp. 142–143.

Default Sensory Hedonism (DSH)

i. Every episode of sensory pleasure is intrinsically good; every episode of sensory pain is intrinsically bad.
ii. The intrinsic value of an episode of sensory pleasure is equal to the number of hedons of pleasure contained in that episode; the intrinsic value of an episode of sensory pain is equal to –(the number of dolors of pain contained in that episode).
iii. The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of sensory pleasure and pain contained in the life, in such a way that one life is intrinsically better than another if and only if the net amount of sensory pleasure in the one is greater than the net amount of sensory pleasure in the other.

2.2 Some Merits of Attitudinal Hedonism

Attitudinal hedonism has attractive advantages over the default theory in that the simple theory or modified versions of this theory escape troubling objections that have been advanced against the default theory. I limit discussion to three objections that many have considered to be of the most serious sort that afflict versions of sensory hedonism akin to the default theory. For each objection, I first summarize the objection. I then explain what resources the simple theory or an adjusted version of the simple theory has at its disposal to circumvent the objection.

2.2.1 A Problem Concerning Nonexistent Pleasures

It certainly seems possible that some lives might be good even though they do not contain any sensory pleasures. Imagine a person who strives to achieve peace and tranquility in her life. Her ideal of a good life for her is a life given up to meditation, a life devoid of sensory pleasure and pain. Suppose this person achieves this ideal. She is completely satisfied with various facts about her life, including the fact that she is not experiencing any pleasure or pain. The default theory implies that this person did not have a good life; her life was low in welfare value. But this seems wrong. Our ascetic enjoyed her life; she got everything she wanted at every moment. The simple theory does not have this disturbing implication. Since the ascetic takes intrinsic attitudinal pleasure in the fact that she gets peace and quiet and (we may assume) takes no displeasure in this fact, SIAH yields the result that her life is high in welfare value (Feldman 2004, pp. 49–51, 68–69).

2.2.2 A Problem Concerning False Pleasures

One of the most forceful objections to default sensory hedonism turns on the fact that if two lives are alike with respect to sensory pleasures and pains, then those lives are of equal (welfare) value. This is true even if the pleasures in the one are "true pleasures"; they all derive from, or are connected with, facts that obtain in

the world, whereas the pleasures in the other are "false pleasures"; they all are associated with what is perceived or taken to be true but with what is not in fact true. As an illustration, imagine that a first businessman experiences many episodes of sensory pleasure, each resulting from something he takes to be true, such as being respected by his colleagues, but that is in fact false. His associates actually think little of him. If he were to discover the truth, this first businessman would be filled with sensory pain. The second businessman, a cousin of the first, enjoys just the sorts of thing that the first does. But he is not relevantly deceived: the objects of his sensory pleasures (and pains) are all "real." We may suppose that the two lives are phenomenologically indiscernible from the "inside." The default theory has the counterintuitive implication (as some see it) that the value of the life for the deceived businessman is just the same as the value of the life for his cousin.

Attitudinal hedonism, in its simple form, succumbs to a similar objection. We may suppose that a person's life contains far more intrinsic attitudinal pleasures than intrinsic displeasures, but each of the states of affairs in which the person takes intrinsic pleasure is false. Again, some may object that this person's life should not be, contrary to what the simple theory implies, rated as highly intrinsically good (see, e.g., Griffin 1986, pp. 8–9; Sumner 1996, pp. 94–98; Kagan 1998, pp. 34–36; Adams 1999, p. 84; and Feldman 2004, pp. 109–114). Amplifying this sort of objection, imagine that the first businessman takes intrinsic attitudinal pleasure in many states of affairs, believing that they are all true. But they are in fact all false. The second businessman takes intrinsic pleasure in things just like the ones the first businessman enjoys. Further, he is not relevantly deceived: the objects of his attitudinal pleasures (and displeasures) are all as he takes them to be. Stipulate that the two lives "feel" exactly the same from the "inside" (Feldman 2004, p. 110; Kagan 1998, pp. 34–36). The simple theory implies that these lives are of equal value. To many, this may seem unacceptable. So, some may conclude that the simple theory is refuted.

Feldman proposes that if we think that this is a legitimate shortcoming of SIAH, this theory, unlike the default theory, may be tailored to accommodate the suggestion, implicit in the objection, that the value of an episode of intrinsic attitudinal pleasure be appropriately sensitive to truth. *Truth-adjusted intrinsic attitudinal hedonism* (TIAH) adjusts the values of episodes of intrinsic attitudinal pleasure according to whether the pleasures are taken in false objects but not of intrinsic displeasures for reasons that Feldman discusses (2004, p. 111). On this adjustment,

> [W]e can say that the fundamental goods are takings of intrinsic attitudinal pleasure in various states of affairs, but we can modify this by saying that such takings of pleasure enhance the value of a life more when they are takings of pleasure in true states of affairs. This single modification yields a view according to which the life of the deceived businessman is much less good in itself for him, even if internally indiscernible from the life of his cousin, the undeceived businessman, whose mental life feels to him just like his cousin's feels to his cousin, but whose family and colleagues are in reality as they appear to him to be (Feldman 2004, p. 111).

The suggestion in this paragraph is that the simple theory can be enriched with the supplement that taking intrinsic pleasure in a true state of affairs *augments* the

value of that pleasure—it makes the pleasure better (2004, pp. 112–114, 181–182). This truth-sensitive hedonism has it that truth has no independent welfare value; rather, truth enhances the intrinsic value of the only things that are the bearers of welfare value—atoms of intrinsic attitudinal pleasure and displeasure. On the truth-adjusted theory, basic intrinsic value states are of the form *S takes intrinsic attitudinal pleasure at time t of intensity n1 and duration m1 in state of affairs P when P is true (false)*.

It may be instructive to formulate the theory more precisely. Feldman proposes that the *truth-adjusted amount* of intrinsic attitudinal pleasure in an episode of intrinsic attitudinal pleasure is determined in this way: first, find the intensity and duration of the pleasure in that episode (as we have explained, the duration will be a minimal one on the assumption that time is discretized). The product of these things gives us the "raw amount" of intrinsic attitudinal pleasure in that episode. If the pleasure is taken in a true object, multiply this product by 1. If the pleasure is taken in a false object, multiply the product by 0.1. I suggest that the multiplication by whatever factor be construed as reflecting the *value* of truth on episodes of intrinsic pleasures. The amount of pleasure in an episode of intrinsic pleasure is determined by the intensity (and duration, though durations are all taken to be minimal) of the pleasure. It will be convenient to say that the result of such products is the truth-adjusted amount of intrinsic attitudinal pleasure in that episode. Let the value of an episode of intrinsic pleasure depend upon the truth-adjusted amount of pleasure it contains. The theory can be summarized in this way:

Truth-Adjusted Intrinsic Attitudinal Hedonism (TAIAH)

i. Every episode of intrinsic attitudinal pleasure is intrinsically good; every episode of intrinsic displeasure is intrinsically bad.
ii. The intrinsic value of an episode of intrinsic attitudinal pleasure is equal to the truth-adjusted amount of pleasure contained in that episode; the intrinsic value of an episode of displeasure is equal to –(the amount of intrinsic displeasure contained in that episode).
iii. The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and displeasure contained in the life, in such a way that one life is intrinsically better than another if and only if the net truth-adjusted amount of intrinsic attitudinal pleasure in the one is greater than the net truth-adjusted amount of that sort of pleasure in the other (2004, p. 112).

One might, of course, wonder about why, exactly, attitudinal hedonism should be truth-adjusted: even if one agrees that the second businessman's life is better for him than the first businessman's life is for him (the first), we have not yet specified the grounds on which we should reach this conclusion. Various axiologists have advanced different considerations for truth adjustment, whether they are hedonists, desire-fulfillment theorists, or yet other sorts of welfare theorists. James Griffin, for example, says that in important areas of his life, he would prefer bitter truth to comfortable delusion. Even if he were surrounded by consummate actors who were able to give him "sweet simulcura of love and affection," he would prefer truth. He adds that he would prefer truth "not because it would be morally better, or aesthetically better, or more noble, but because it would make for a better life

for me to live" (Griffin 1986, p. 9). Robert Nozick claims that he would not plug into an "experience machine" because besides the introspectible features of our experiences, we also want to *do* certain things and *be* a certain sort of person (Nozick 1974, p. 43). L. W. Sumner proposes that the trouble with theories such as the simple theory is that according to them, we may assess how well our lives are going just by attending to how they seem from the inside, "bracketing off all questions of their anchoring in the external world" (Sumner 1996, p. 98). He says that any theory with this implication is too "interior and solipsistic to provide a descriptively adequate account of the nature of welfare" (1996, p. 98). Perhaps unifying these seemingly diverse rationales is the thought that if you care about truth—if you take truth to be an important constituent of *your* well-being—then any theory of welfare such as the simple theory, on which your welfare turns entirely on how your life "appears to you from the inside," is defective. To many of us, truth does indeed matter.

Although the truth-adjusted theory deflects the objection from false pleasures, it falls victim to another problem; it is not sensitive to the freedom of intrinsic pleasures (and intrinsic displeasures). Before developing this concern, let us look at another alleged problem that some have judged tells against both the default theory and the simple theory.

2.2.3 A Problem Concerning Base Pleasures

With the default theory in our evaluative sights, suppose again that a person's life contains far more sensory pleasures than pains but this time all his pleasures derive from "worthless" or base objects. Moore speaks of "a perpetual indulgence in bestiality" ([1903] 1962, Chapter III, Section 56). We may take Moore and like-minded people to be objecting that this person's life should not, unlike what the default theory implies, qualify as highly intrinsically good. A similar problem seems to infect the simple theory. Imagine that the person's life contains far more intrinsic attitudinal pleasures than intrinsic displeasures and that all his intrinsic pleasures are pleasures taken in worthless or base objects. The simple theory, just like the default theory, implies that this person's life ranks high in welfare value.

Once again, attitudinal hedonism appears to have the resources to meet this objection. Feldman describes a way in which we could adjust the value of an intrinsic pleasure (or intrinsic displeasure) to reflect the extent to which the object of that pleasure (or displeasure) deserves to be enjoyed:

> [T]he value of a pleasure is enhanced when it is pleasure taken in a pleasure-worthy object, such as something good, or beautiful. The value of a pleasure is mitigated when it is pleasure taken in a pleasure-unworthy object, such as something evil, or ugly. The disvalue of a pain is mitigated (the pain is made less bad) when it is pain taken in an object worthy of pain, such as something evil, or ugly. The value of a pain is enhanced (the pain is made yet worse) when it is pain taken in an object unworthy of this attitude, such as something good or beautiful (2004, p. 120, note omitted).

A basic intrinsic value state on this axiology would be a state of affairs of this form: *S takes intrinsic attitudinal pleasure at time t of intensity n1 and duration m1 in state of affairs P, while P is worthy of pleasure of intensity n2 and duration m2*. An adjustment of the *value* of pleasure in a complex state of affairs of this sort, this time not for truth but for object-worthiness, yields the *desert-adjusted amount* of pleasure it contains. The values of displeasures are similarly attuned. The resulting theory generates, it seems, the desired outcomes in cases involving base or malicious pleasures. The theory runs as follows:

> Object-Worthy Intrinsic Attitudinal Hedonism (OWIAH)
>
> i. Every episode of intrinsic attitudinal pleasure is intrinsically good; every episode of intrinsic displeasure is intrinsically bad.
> ii. The intrinsic value of an episode of intrinsic attitudinal pleasure is equal to the desert-adjusted amount of pleasure contained in that episode; the intrinsic value of an episode of displeasure is equal to –(the desert-adjusted amount of intrinsic displeasure contained in that episode).
> iii. The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and displeasure contained in the life, in such a way that one life is intrinsically better than another if and only if the net deserted-adjusted amount of intrinsic attitudinal pleasure in the one is greater than the net desert-adjusted amount of that sort of pleasure in the other (2004, p. 121).

Just like the truth-sensitive theory, though, I believe that OWIAH is deficient in that it is insufficiently sensitive to the freedom of intrinsic pleasures and displeasures.[4] We turn to the topic of the bearing of freedom on the value of life atoms in the next chapter.

[4] I do not deny that there may be other deficiencies with the theory (see, e.g., Sections 4.3 and 5.5).

Chapter 3
Freedom-Sensitive Versions of Attitudinal Hedonism

3.1 The Freedom of Attitudes

To develop freedom-sensitive versions of intrinsic attitudinal hedonistic theories of welfare, we need to address when attitudinal pleasures (or displeasures) are free and when they are unfree. To do so, I first say something more general about the freedom of attitudes, feelings, and emotions, or abbreviating, I say some general things about the freedom of feeling-states. The freedom at issue, again, is the freedom that moral responsibility requires.

Suppose we are morally responsible for at least some of our intentional actions. Then it is highly credible that we are morally responsible, too, for at least some of our emotion-tokens, particular instances of joy or gratitude, for example; we are also morally responsible for at least some of our feeling-tokens, tokens that are affective but do not qualify as emotions; and we are morally responsible, as well, for, minimally, some of our attitude-tokens, such as tokens of our taking pleasure in various facts. If we are responsible for such feeling states, then since responsibility presupposes freedom or control, we must have responsibility-relevant control with respect to these states. Commenting on such control Alfred Mele writes

> That we have some control over what we feel and over the intensity of our emotions and other feelings is clear. We stem a discomforting flow of sympathy for a character in a film by reminding ourselves that he is *only* a character. . . . The woman who regards her anger at her child as destructive may dissolve or attenuate it by forcing herself to focus her attention on a cherished moment with the child. The timid employee who believes that he can muster the courage to demand a raise only if he becomes angry at his boss may deliberately make himself angry by vividly representing the injustices that he has suffered at the office. . . . These are instances of what I call *internal* control. . . . Many emotions and feelings are subject to external control as well—control through one's overt behavior. Jill knows that if, for some reason, she wants to be angry, a phone call to her mother will turn the trick. Jack defeats mild depression by calling his sister (1995, p. 106).[1]

We have indirect responsibility-relevant freedom (or control) over something only if we have control over it by virtue of having control over something else.

[1] See also Oakley (1992, pp. 122–159).

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

We have direct control over something only if our control over it is not indirect. Similarly, we are indirectly responsible for something only if we are responsible for it via being responsible for something else; we are directly responsible for something only if we are not indirectly responsible for it. Responsibility tracks control in that we can be directly responsible for something only if it is in our direct control and if something is in our indirect control we can, at best, be indirectly responsible for it (provided that the thing is not also in our direct control). It is unlikely that the control we or most of us have over our feeling-tokens over which we do have control is direct; it is far more tenable that this control is indirect. On pain of avoiding an infinite regress, all indirect control in the end must trace to something over which we have direct control.

Assume that if we have direct control over certain things, we have direct control only over our decisions. (This is not essential to what follows; should one disagree with this assumption, simply supply one's favorite candidate—some "basic action"—as the candidate for whatever it is over which we have direct control. The freedom of decisions, then, should trace to the direct freedom of these basic actions.) If we have direct control solely over our decisions, then the indirect control (or freedom) that we have over our attitudes will, in the end, derive from the direct control (or freedom) that we have over our decisions. I assume, here and in what follows in this chapter, that each person's life contains at least some episodes of intrinsic attitudinal pleasure and some of intrinsic displeasure. To the extent that our decisions are free, to that extent our pertinent attitudes will be free (supposing nondeviant, "smooth" production of the attitude *via* the decision).

Consider another example of Mele's:

> For years, Tom has felt inferior to his friend and colleague, Betty owing to her greater success as a scholar. Tom, who does not wish Betty ill, has just learned that a paper of Betty's recently received its fifth rejection, and he finds that he is pleased by the news. Tom judges that his pleasure is unwarranted by the evoking stimuli and that his no longer being pleased by Betty's failure would be best, all things considered. Moreover, he believes that he can eradicate the pleasure simply by absorbing himself in his work. Nevertheless, he chooses to savor the moment, to allow himself to enjoy the news—but without revising his decisive better judgment. . . . Though it may often happen that our choices are in line with our better judgments, sometimes we choose or intend to do one thing while judging something else better. . . . Notice also that insufficient effort need not be due either to a judgment that taking preventative or ameliorative measures is not worth the trouble or to a misjudgment about the amount (or kind) of effort required. Tom may correctly judge that his absorbing himself in his work would put an end to his feeling-state, and he may be able to do this. But if he is less than perfectly self-controlled, he may make a half-hearted, unsuccessful attempt (1995, p. 107).

We may assume that Tom takes intrinsic attitudinal pleasure in the fact that Betty's paper has been rejected for a fifth time. He judges that it is better (from his perspective of what is better) that he not enjoy such pleasure, but he persists in being pleased. Mele says that Tom believes (correctly) that he can eradicate the pleasure by absorbing himself in his work. But Tom chooses to savor the moment, and he continues to take intrinsic pleasure in Betty's failure. Assume that if Tom were to have eradicated the pleasure, he would have done so, at least partly, by deciding to

absorb himself in his work. Similarly, we may suppose that Tom persists in taking intrinsic pleasure in Betty's failure partly via the decision to prolong "savoring the moment." Since it is true that Tom *could* have put an end to his taking pleasure in Betty's failure by deciding to absorb himself in work, but prolongs the pleasure by making some other decision instead, it is highly plausible that Tom's being pleased at Betty's failure is an atom of pleasure that is free. Given his decisive better judgment, we may further suppose that Tom's believes, at the time that he takes pleasure in Betty's failure, that he is doing moral wrong in taking such pleasure. It would then not be untoward to conclude both that Tom is morally blameworthy for taking such pleasure and that this episode of pleasure is akratic.

Provisionally, we may say that an episode (or atom) of intrinsic attitudinal pleasure is free if you freely take intrinsic pleasure in some object. Just as a desire may be "engineered" into you against your will, so you may be made to adopt or assume certain attitudes against your will. Such engineered-in attitudes would be "inauthentic," "not truly your own," or simply "not free." Construe "free" in the last sentence in whatever way in which this word is construed when it is said that responsibility requires that one's actions be free.

3.2 Neo's Case

Now ponder the following case that strongly suggests that freedom affects the intrinsic value of episodes of pleasures and displeasures that are pertinent to well-being. In a scenario, much like the one depicted in the movie hit, *The Matrix*, Neo is kept alive in a bed of nutrients. In this scenario ("Matrix-1"), Neo's brain is connected to a sophisticated computer. When the computer is appropriately programmed, suitable centers in Neo's brain are stimulated. As a result, all the while lying in his slimy, life-sustaining womb, Neo lives a make-believe life. He is completely in the dark about living such a life. Further, he is happy in this fabricated life: he is made to take intrinsic attitudinal pleasure in many different states of affairs, but none of these states of affairs is true. We may also suppose that his artificial life includes few, widely scattered episodes of intrinsic displeasure. Since Neo's life contains a healthy balance of net attitudinal pleasure, the simple theory implies that Neo's life is pretty good in itself for Neo. But this implication of the simple theory is highly troubling. How can imprisonment in a slimy womb be good for anyone? Neo may be *happy*, but it seems farfetched to say that his *life* is going well for him. To drive home this point, we might invoke a counterpart of Neo, Neo*, who is not a victim of the Matrix. Neo* lives the very same sort of life as Neo with the exception that his life is "real"; the objects of Neo*'s intrinsic attitudinal pleasures and intrinsic displeasures are all just as he takes them to be. The simple theory implies that the two lives are of equal value. But this is dubious. So the simple theory is to be rejected.

Some people, it seems, have suggested that examples such as Matrix-1 refute sensory hedonism (see, e.g., Nozick 1974, pp. 42–45; Griffin 1986, Chapter 1; Sumner 1996, pp. 96–98). Maybe others will say that the examples refute all forms

of attitudinal hedonism. On my view, in contrast, Matrix-1 gives us powerful reason to revise the simple theory. The underlying thought is that perhaps freedom should be a constituent of life atoms.

A slightly modified scenario, "Matrix-2," suggests that truth-adjusted intrinsic attitudinal hedonism is also deficient for a reason similar to the one just suggested regarding the inadequacy of the simple theory. Assume that in Matrix-2, Neo is made to take intrinsic attitudinal pleasure in states of affairs each of which is true. (So, for instance, Neo is made to take pleasure in *Bush's being president of the U.S. in 2006*.) Neo's life, then, is full of episodes of intrinsic attitudinal pleasure in true states of affairs. Assume, in addition, that Neo does not take pleasure in any false state of affairs and that he does not take intrinsic displeasure in any state of affairs. I doubt whether we would want to judge that Neo's *life* is highly intrinsically good for Neo, contrary to what both the simple theory and the truth-adjusted variant imply. Surely, spending one's entire life in a slimy womb at the mercy of evil manipulators cannot be anyone's vision of the good life. How could such a life, even if it is full of engineered-in episodes of intrinsic attitudinal pleasure in true states of affairs and devoid of episodes of intrinsic displeasures, be a life good in itself for the one who lives it? Again, we need not (though some may) dispute that Neo in Matrix-2 might well be *happy*; our concern is not with happiness but with *welfare*.

What more, initially, may be said in favor of the view that it is at least prima facie plausible that whether intrinsic attitudinal pleasures and intrinsic displeasures are free or not free affects welfare? First, if it is *plausible* (as I think it is) that takings of pleasure enhance the value of a life more when they are takings of pleasure in true states of affairs, so it is plausible that takings of pleasure enhance the value of a life more when they are free takings of pleasure in true states of affairs. Suppose you are manipulated (in the absence of your knowledge of being manipulated) into taking intrinsic attitudinal pleasures in various objects. The crafty neuroscientists want to see how far they can go in controlling various attitudes of others; they are particularly interested in experimenting with attitudinal pleasures and displeasures. Even if the objects in which you are made to take intrinsic attitudinal pleasure are true, it still seems that the relevant episodes of attitudinal pleasure are not as intrinsically good as they would be if they were free episodes.

Second, it is credible to suppose that the intrinsic value of an action's (total) outcome depends at least partly on the fit between what pleasure people in the outcome get and what pleasure they deserve to get, the better the fit the greater the intrinsic value (see, e.g., Feldman 1997, pp. 151–174). Responsibility, roughly, is a desert base in this sense: if one is morally praiseworthy, and in virtue of being praiseworthy, deserves pleasure but fails to get the pleasure one deserves, this is not so intrinsically good and perhaps intrinsically bad. If one is not free and gets pleasure that is not deserved, again this is not as intrinsically good as one's getting pleasure that is deserved because one is morally praiseworthy. So we have some reason to believe that freedom affects the intrinsic value of a basic. (Of course, these considerations leave it open whether desert, and, thus, freedom affect the value solely of world atoms and not of life atoms.)

Some people may insist that it is not so bad to take intrinsic pleasure in false states of affairs. Maybe, as Feldman is willing to entertain, takings of pleasure enhance the intrinsic value of a life *more* when they are takings of pleasure in true states of affairs. Reconsider Neo's predicament. Neo is subject to a sort of manipulation, which is such that he is made to take intrinsic attitudinal pleasure in various objects. In the case as originally presented, we assumed that all these objects are false. We said that the life Neo leads is an unattractive life; it is low in welfare value. Even if such a life were intrinsically good for Neo, it would, at best, be minimally intrinsically good for him. What pertinent factor accounts for this negative judgment?

It appears that the relevant explanatory factor cannot be merely that Neo takes intrinsic attitudinal pleasure in false states of affairs. We are assuming that taking pleasure in false states of affairs does enhance the intrinsic value of a life but not to the extent that taking pleasure in true states of affairs does. If taking pleasure in false states of affairs is the relevant explanatory factor, then Neo's life in his slimy womb would be highly intrinsically good for Neo provided that his make-believe life is choke full of episodes of intrinsic attitudinal pleasure and devoid of episodes of intrinsic displeasure (though less good than it would be were Neo to take intrinsic pleasure in a suitable number of true states of affairs). But it is false that Neo's life in his slimy womb would be highly intrinsically good for Neo. It follows that it is false that taking pleasure in false states of affairs is the pertinent explanatory factor. The relevant explanatory factor seems to be that Neo is not free; in particular, his value-conferring attitudes are not free. Matrix-2 supports this diagnosis.

3.3 In Support of Freedom-Sensitive Attitudinal Hedonism

If one believes that freedom makes a life intrinsically better for the person who leads it, just as one may believe that truth makes a life intrinsically better for the person who lives it, then, as I illustrate below, this can be explained by freedom-sensitive versions of intrinsic attitudinal hedonism. My proposal, though, is stronger. I think that free intrinsic pleasures *do* enhance well-being, more so than otherwise similar pleasures that are not free. I believe that Neo's case provides strong, prima facie support for this view. I may, of course, be challenged to produce *an argument* for the view at issue. Some comments regarding this strong demand are in order.

As a preliminary remark, suppose one believes that taking intrinsic attitudinal pleasure in true states of affairs enhances well-being, or that deserved pleasures enhance well-being, or that pleasure in the good is itself good, or that a pleasure in an object that deserves to have pleasure taken in it is better than an otherwise similar pleasure in an object that does not deserve to have pleasure taken in it, or that the fulfilling of a moral obligation is good. I think each of these positions is plausible. I confess to not having come across a single "direct" argument for any of these views: I know of no argument that proceeds from premises more evident than the claim in question to the conclusion that, for instance, truth enhances the value of an episode of intrinsic attitudinal pleasure. Rather, in support of each of these positions, typically we are presented with a scenario such as the one involving the

two businessmen or one in which a person takes pleasure in wallowing with pigs. Then we are invited to reflect on what factor or factors the scenarios suggest are pertinent to welfare. So in this respect, the view that freedom enhances the value of intrinsic attitudinal pleasures—other things equal, free pleasures are better than otherwise similar pleasures that are not free—is not any worse off.

Second, in defense of the view that free pleasures are better, one should keep in mind an important distinction. As I have emphasized, we should distinguish between Neo's *being happy* in Matrix-2, or Neo's being made to feel sensory pleasure or being made to take intrinsic attitudinal pleasures in various objects that are true, and Neo's *welfare*. Even if one is a hedonist of sorts about value and welfare, I see no inconsistency in the view that Neo may well be happy in Matrix-2 but that his welfare level in the slimy womb is very low. To account for his low welfare level, an appeal to the (un)freedom of his intrinsic attitudinal pleasures seems unavoidable.[2]

Third, in support of the relevance of freedom to welfare, one reflects carefully on cases such as Neo's. One moves back and forth between examples of this sort and theoretical considerations until one arrives at what looks like an overall tenable position. In this respect, once again I am no worse off than a proponent of the claim that pleasure in true states of affairs enhances well-being, or that a pleasure in an object that deserves to have pleasure taken in it is better than an otherwise similar pleasure in an object that does not deserve to have pleasure taken in it. Surely, it is upon such considerations of reflective equilibrium that proponents of these other positions partly rely to support their positions.

Finally, one can build up a negative defense of one's axiological view. One can consider objections against one's view and deflect them. I turn to this negative defense in Section 3.6 below.

3.4 Varieties of Freedom-Sensitive Attitudinal Hedonism

Before defending further the suggestion that freedom has a bearing on life atoms, it will be helpful to examine a number of variants of freedom-sensitive attitudinal hedonism. For each, I first formulate the theory. I then discuss some of its merits and shortcomings and trace a sample of its implications for well-being. Then I say something about what I think is the most promising of these contenders.

3.4.1 Robust Freedom-Sensitive Intrinsic Attitudinal Hedonism

The first variation has fairly robust commitments to freedom. On this theory, the life atoms all look relevantly like this:

---

[2] Sumner claims that welfare consists in *authentic* happiness, the happiness of an informed and autonomous subject (1996, p. 139).

Life Atom-1: At noon on Tuesday, October 16, 2001, Bob freely takes intrinsic attitudinal pleasure of intensity +8 in the fact that Bob's beer is frosty cold.

There are corresponding atoms concerning intrinsic displeasure. The intuitive idea of this form of hedonism is that it is only free episodes of intrinsic attitudinal pleasures and displeasures that have intrinsic value; unfree pleasures and unfree displeasures are intrinsically worthless. The theory runs as follows:

Robust Freedom-Sensitive Intrinsic Attitudinal Hedonism

i. Every free episode of intrinsic attitudinal pleasure is intrinsically good; every free episode of intrinsic displeasure is intrinsically bad.
ii. The intrinsic value of a free episode of intrinsic attitudinal pleasure is equal to the amount of pleasure contained in that episode; the intrinsic value of a free episode of displeasure is equal to –(the amount of displeasure contained in that episode).
iii. The intrinsic value of a life (for a person) is entirely determined by the intrinsic values of the free episodes of intrinsic attitudinal pleasure and displeasure contained in the life, in such a way that one life is intrinsically better than another if and only if the net amount of (free) intrinsic attitudinal pleasure in the one is greater than the net amount of (free) intrinsic attitudinal pleasure in the other.

The theory can be combined with other considerations thought pertinent to the intrinsic value of lives. For instance, as previously mentioned, if we believe that takings of pleasure enhance the value of a life more when they are takings of intrinsic pleasure in *true* states of affairs, then we might want to adjust the values of intrinsic pleasures for both freedom and truth. The second clause of the robust theory could, consequently, be modified to accommodate truth in this way: *the intrinsic value of a free episode of intrinsic attitudinal pleasure is equal to the truth-adjusted amount of pleasure contained in that episode*. Or we may want to adjust the values of intrinsic pleasures for freedom and the extent to which the objects in which one takes pleasure deserve to have pleasure taken in them. However, since my concern is primarily with freedom, I steer clear of these more complex forms of intrinsic attitudinal hedonism.

An advantage of the robust view is that it yields an intuitively correct rating of Neo's life in Matrix-1 and Matrix-2. Neo takes pleasure in many things but all these pleasures are unfree. In addition, he takes little or no displeasure in anything. Thus, the robust view implies that his life is not good in itself for him; it is intrinsically worthless.

Some, though, may think that a serious shortcoming of the robust theory is that it has unacceptable implications concerning the intrinsic value of unfree episodes of pleasure and displeasure. They might claim that, regarding life atoms, maybe it is plausible to suppose that unfree intrinsic attitudinal pleasures are less good than free intrinsic pleasures, as the robust theory implies. But the robust theory implies something more about the value of unfree pleasures: it implies that they are intrinsically worthless. Many may find this troubling. It may further be added that it is *not* reasonable to theorize that unfree intrinsic displeasures, in comparison to otherwise similar free displeasures, are intrinsically worthless. This, it may be submitted, is confirmed by Matrix-3. Matrix-3 is just like Matrix-1, save that Neo does not take pleasure in any state of affairs and that he is made to take intrinsic

displeasure in states of affairs all of which are true. Let Leo be a counterpart of Neo's whose life is internally indiscernible from Neo's (in Matrix-3) but a life that is "real." The objection is that if Leo's life is bad in itself for Leo, Neo's life should be *just* as bad in itself for Neo even though all of Neo's displeasures are unfree. Hence, it is false that unfree intrinsic displeasures are intrinsically worthless.

There is a second objection against the robust theory. The objection is that freedom has a less marked effect on intrinsic value than this theory implies. The intrinsic value of an episode of intrinsic attitudinal pleasure is *enhanced* if that episode is free. And perhaps intrinsic displeasures that are not free are either *less* intrinsically bad or perhaps *worse* than ones that are free.

If one accepts these objections, one may well be drawn to more modest versions of freedom-sensitive attitudinal hedonism.

3.4.2 Asymmetric Freedom-Sensitive Intrinsic Attitudinal Hedonism

On the first modest version, freedom affects the intrinsic value of episodes of intrinsic attitudinal pleasure—it enhances the value of pleasure—but has no effect on the intrinsic value of episodes of intrinsic displeasure. What we may call the "freedom-adjusted" amount of intrinsic attitudinal pleasure in an episode of intrinsic attitudinal pleasure is determined in this way: first, find the raw amount of intrinsic pleasure in the episode by multiplying the intensity and duration of the pleasure in that episode. If the episode of pleasure is free, multiply the product by 1. If the episode of pleasure is unfree, multiply the product by 0.1. Again, the multiplication should be taken to reflect the value of freedom on episodes of pleasures. Freedom has no independent welfare value, but it enhances the intrinsic value of positive life-atoms, episodes of intrinsic attitudinal pleasure. We may say that the result of the multiplication is the *freedom-adjusted amount* of intrinsic attitudinal pleasure in that episode. Let the value of a pleasure depend upon the freedom-adjusted amount of pleasure it contains. The theory may be formulated in this fashion:

Asymmetric Freedom-Sensitive Intrinsic Attitudinal Hedonism

i. Every episode of intrinsic attitudinal pleasure is intrinsically good; every episode of intrinsic attitudinal displeasure is intrinsically bad.
ii. The intrinsic value of an episode of intrinsic attitudinal pleasure is equal to the freedom-adjusted amount of pleasure contained in that episode; the intrinsic value of an episode of displeasure is equal to –(the amount of displeasure contained in that episode).
iii. The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and displeasure contained in that life in such a way that one life is intrinsically better than another if and only if the net freedom-adjusted amount of intrinsic attitudinal pleasure in the one is greater than the net freedom-adjusted amount of that sort of pleasure in the other.

On the credit side of this theory, some will approve of its implication that Neo's life in Matrix-3 is intrinsically bad. On the debit side, though, some may be convinced that lack of freedom mitigates but does not eradicate the badness of

an episode of displeasure. For this reason, they may opt for *symmetric freedom-sensitive intrinsic attitudinal hedonism*.

3.4.3 Symmetric Freedom-Sensitive Intrinsic Attitudinal Hedonism

According to one type of moderate theory that is symmetric, adjustments, in the fashion described when formulating the asymmetric theory, are made to the values of episodes of intrinsic attitudinal pleasures and like adjustments are made to the values of episodes of intrinsic displeasures. On this theory, unfree pleasures are good but not as good as free ones *and* unfree displeasures are bad but not as bad as free ones. As some will see it, in this theory's favor is its implication that Neo's life in Matrix-2 is intrinsically bad. Working against it, though, is the fact that, if true, then Neo's life in Matrix-1 is intrinsically good and may be highly so. This may strike some as hard to swallow but it may have to be swallowed.

There is yet another sort of symmetric theory. It states, like the first, that unfree pleasures are good though not as good as free ones. It differs from the first in its contention that unfree displeasures are *worse* than free displeasures. It shares with the first variety what some may regard as the troubling implication that, if true, Neo's life in Matrix-1 (and in Matrix-2) is highly intrinsically good for Neo. Again, this is an implication with which we may have to live.

It may be worthwhile saying a few things about this implication. I said that Matrix-1 and Matrix-2 give us reason to believe that whether or not episodes of intrinsic attitudinal pleasure and intrinsic displeasure are free has a bearing on welfare. Neo may be happy in the Matrix but his life is not going well for him; his life has low welfare value. On the assumption, though, that his life in the slimy womb is replete with episodes of intrinsic attitudinal pleasure and devoid of episodes of intrinsic displeasure, either symmetric theory implies that Neo's life in the Matrix may he highly good for him. Does this implication, then, not call into question the initial thought that the freedom of attitudinal pleasures and displeasures can affect welfare? The answer is that the implication leaves the initial thought unblemished. To see this, we need to keep in mind the following. Compare the life Neo lives in, for instance, Matrix-2 with the life that is otherwise similar but "real." In the first life, Neo takes intrinsic attitudinal pleasure in states of affairs, each of which is true, each of his attitudinal pleasures is unfree, and he does not take intrinsic displeasure in any state of affairs. In the second life, again Neo takes intrinsic pleasure in true states of affairs, he does not take displeasure in any state of affairs, but all his pleasures are free. The simple theory (SIAH) and the truth-adjusted theory (TAIAH) yield the result that the two lives are of equal value; they imply that each is highly good in itself for Neo. In contrast, either symmetric theory implies that though Neo's life in the Matrix may be highly good for Neo, the "real" life in which his pleasures are all free is even *better*. Thus, the freedom of pleasures has a definite impact on well-being.

It is interesting and challenging to argue in favor of one or the other of the second components concerning displeasure of the two sorts of symmetric theory.

On the first, free intrinsic displeasures are worse than unfree ones; on the second, free intrinsic displeasures are better than (they are not as bad as) unfree ones. It is advisable to look more closely at these components.

Regarding the first symmetric theory, is it *credible* that free displeasures are worse than otherwise similar unfree ones? Well, perhaps so. Since it is intrinsic displeasures—intrinsic attitudinal pains—that are of concern, confine attention to displeasures that are divorced from sensory pains. When one freely takes displeasure in something, we can say that one's agency is "invested" in the attitude; the attitude causally issues from agential elements, such as desires, beliefs, and values that are "truly the agent's own." The pertinent attitude is under one's indirect control. With an unfree episode of displeasure, such as Neo's (in Matrix-3), Neo's agency is not pertinently implicated in the episode. The causal process (or significant parts of that process) that generates the attitude bypasses Neo's capacities of deliberative control. And because of such bypassing, we may say that Neo's agency is not invested in the relevant displeasures. It may be proposed that when agency is invested in an episode of displeasure, this episode is intrinsically worse for the agent than an otherwise similar attitude in which agency is not invested. We can summarize the suggestion in this way: displeasures in which agency is invested are displeasures that are one's own; displeasures in which agency is not invested are displeasures that are not one's own. Displeasures that are one's own are intrinsically worse for one than are displeasures that are not one's own. Maybe analogous things are true about attitudinal pleasures. Attitudinal pleasures in which agency is invested are intrinsically better for one than are attitudinal pleasures in which agency is not invested.

This position on displeasures *somewhat* resembles a position concerning another sort of normative appraisal, this time an appraisal of agents. If one freely believes that one is doing something that is morally wrong and one intentionally does it partly on the basis of this belief, then (provided other conditions of moral responsibility are met), one is morally blameworthy for the thing; one expresses what one "stands for" in one's actions. One is "off the hook," though, if, say, one's belief or action is not free. The parallelism with the axiological appraisal should be fairly obvious. If one's episode of displeasure is free, we may think of the attitude as expressing something about the moral sensitivity of the agent. If, though, one's attitude is not free, the attitude fails to express anything about the agent's morally significant attitudes.

This line of reasoning, however, is not conclusive. It assumes that agency investment enhances the value of an intrinsic displeasure; the investment makes the displeasure worse. But it may be thought that when an agent is invested in her attitudinal pleasures and displeasures—when her agency is implicated in her episodes of pleasure and displeasure—this is a good thing. In virtue of agency investment's being good, freedom enhances the value of an intrinsic attitudinal pleasure. Further, freedom has the effect of making something that is bad—a displeasure—less bad; freedom mitigates the value of an intrinsic displeasure. I confess to being drawn more by this second view—the view that implies that free intrinsic displeasures are not as bad as otherwise similar displeasures that are unfree—than to being drawn

by the first. (We may dub this second view as "favored symmetric freedom-sensitive hedonism.") But I have no decisive argument for my preference.

3.4.4 Pain-Adjusted Freedom-Sensitive Intrinsic Attitudinal Hedonism

Some, who cannot live with the implication of either of the symmetric views that Neo's life in Matrix-1 is highly intrinsically good for Neo may opt for yet another variation of freedom-sensitive hedonism. This variation, *pain-adjusted freedom-sensitive intrinsic attitudinal hedonism,* adjusts the intrinsic values of episodes of intrinsic displeasure according to whether or not the episodes are free and it declares, as intrinsically worthless, episodes of intrinsic attitudinal pleasure that are unfree. Again, there are two variations of this theory depending upon whether unfree intrinsic displeasures are not as bad as free ones (first version) or whether unfree intrinsic displeasures are worse than free ones (second version). Either variation implies that it is false that Neo's life in Matrix-1 is intrinsically good. I find this view unappealing. If lack of freedom affects the value of a displeasure, it is hard to see why lack of freedom should not also affect the value of a pleasure.

I do not, in this book, undertake the daunting task to argue for the overall superiority of one of these theories. (To put my cards on the table, I am partial to the second symmetric theory.) Rather, paying close attention to whether the agent's attitudinal pleasures and displeasures are not free, I expose pertinent implications of these theories regarding well-being. After all, I want to debunk the idea that freedom leaves unscathed various sorts of axiological appraisal such as evaluations of welfare and assessments of worlds.

3.5 Freedom and Well-Being

The gist of the robustly freedom-sensitive theory is that unfree episodes of intrinsic attitudinal pleasures and intrinsic displeasures are intrinsically valueless. If this robust theory is true, the intrinsic value of lives that are bereft of freedom is undermined: no agent's life will be good in itself for that agent.

With the asymmetric theory, unfree intrinsic attitudinal pleasures are less intrinsically good than free intrinsic attitudinal pleasures, but freedom does not have any impact on intrinsic displeasures; an unfree episode of displeasure is just as intrinsically bad as it would be if it were free. Assuming that the agent's life contains some episodes of intrinsic attitudinal pleasure and some of intrinsic displeasure, and that the former episodes are unfree, the effect of this freedom-sensitive asymmetric theory will be to diminish the overall intrinsic value of the agent's life. If we compare two lives, *life-1* and *life-2*, that contain exactly the same raw amounts of intrinsic attitudinal pleasure and intrinsic displeasure but

differ in that all the episodes of attitudinal pleasure in *life-1* are free, whereas all of the pleasures in *life-2* are unfree, the asymmetric view implies that *life-1* is intrinsically better for its agent than *life-2* is for its agent. This is morally significant.

With either version of the freedom-sensitive theories that are symmetric, the values of both episodes of intrinsic attitudinal pleasure and intrinsic displeasure are adjusted according to whether the episodes are free or unfree. On both theories, free attitudinal pleasures are better than unfree pleasures. On the first, free displeasures are worse than otherwise similar unfree displeasures; on the second, free displeasures, though bad, are less bad than unfree ones. The implications of the symmetric theories for the values of lives are not so clear-cut. Consider three sorts of case. In the first, the agent's life contains more intrinsic attitudinal pleasure than intrinsic displeasure. If either symmetric theory is true, such a life (in which the episodes of attitudinal pleasure and displeasure are unfree—call such a life an "unfree life") will be less overall good than a corresponding life in which the type- or near type-identical episodes of pleasure and displeasure are free (call this sort of life a "free life"). In the second, the agent's life contains much more intrinsic displeasure than intrinsic attitudinal pleasure. If either theory is true, such an unfree life will be overall better than a corresponding free life. In the third, the agent's life contains roughly equal quantities of freedom-adjusted intrinsic attitudinal pleasure and similarly adjusted intrinsic displeasure. If either symmetric theory were true, perhaps such lives would be just as intrinsically good for their agents as they would be if they contained type- or near type-identical episodes of attitudinal pleasures and displeasures that were free.

Finally, *pain-adjusted freedom-sensitive intrinsic attitudinal hedonism* adjusts the values of episodes of intrinsic displeasure according to whether or not they are free and rates as intrinsically worthless unfree episodes of intrinsic attitudinal pleasure. Recall the two versions of this theory: unfree displeasures are not as bad as free displeasures (first version); unfree displeasures are worse than free displeasures (second version). If the first version is true, unfree lives with some intrinsic attitudinal pleasures and some intrinsic displeasures will be overall less intrinsically good than corresponding free lives as will unfree lives with some intrinsic attitudinal pleasure but no intrinsic displeasure. In addition, unfree lives with some displeasure but no attitudinal pleasure will be overall more good than corresponding free lives. If the second version is true, unfree lives with some intrinsic attitudinal pleasures and intrinsic displeasures will, again, be overall less intrinsically good than corresponding free lives as will unfree lives with some intrinsic attitudinal pleasure but no intrinsic displeasure. Unfree lives with some intrinsic displeasure but no intrinsic attitudinal pleasure will be overall worse than corresponding free lives.

An important generalization may now be recorded: assume that each person's life contains some episodes of raw intrinsic attitudinal pleasure and some of raw intrinsic displeasure. Some lives may contain more pleasure than displeasure, others less pleasure than displeasure, and yet others equal amounts of pleasure and displeasure. Given this assumption, regardless of which version of freedom-sensitive hedonism

that we have introduced is presupposed, *free intrinsic pleasures enhance the value of a life, no matter what the total value of intrinsic displeasure in that life* (and, hence, no matter whether the episodes of displeasure in that life are free or not free). If free pleasures enhance well-being in this manner, it would be a good thing for us if the episodes of intrinsic attitudinal pleasures in our lives were free.

3.6 Objections and Replies

I now defend the view that whether or not an episode of intrinsic attitudinal pleasure or intrinsic displeasure is free affects welfare against various objections.

3.6.1 Welfare-Relevant Versus World-Relevant Factors

The first objection is an objection that, if cogent, tells equally against the truth-adjusted theory as it does against the freedom-sensitive versions of attitudinal hedonism that adjust the values of episodes of attitudinal pleasure and displeasure to reflect the extent to which these episodes are free. The objection is that the deceived life or the unfree life (i.e., the life in which the agent's episodes of intrinsic pleasure and intrinsic displeasure are unfree) still rank high in terms of *welfare*; these lives *are* intrinsically good for their agents. But we (mistakenly) judge them unfavorably because factors such as deception and unfreedom contribute negatively to the value of *worlds*. In short, we confuse what we may call "world-relevant factors"—factors that affect the intrinsic value of worlds with "welfare-relevant factors"—factors that bear on the intrinsic value of lives for agents.[3]

Why precisely, though, is it supposed that factors such as deception, truth, falsity, or lack of freedom are world-relevant but welfare-irrelevant? A preliminary thought maybe is that "what you don't know can't hurt you," and therefore truth is not welfare-relevant. Similarly, if all your episodes of intrinsic attitudinal pleasure or intrinsic displeasure are unfree but you don't know that they are unfree, this can't hurt you. You would not feel any different. Recall that Neo's life (in Matrix-2) is internally indiscernible from Neo*'s life, the latter's life being just like Neo's save that Neo*'s life is not a make-believe life.

However, the fact that you would not feel any different if you or your life or a constituent of your life lacked a certain factor provides no definitive reason to believe that the factor is welfare-irrelevant. We may develop this point in connection with commenting on a dispute between those who believe that attitudinal pleasures are not value-conferring unless they are associated with sensory pleasures and those who deny this. Michael Zimmerman distinguishes between being thinly pleased or displeased about something and being thickly pleased or displeased about a thing (Zimmerman 2001, pp. 195–197, 2007, p. 427). To be

[3] On different dimensions or standpoints of value, see, for example, Sumner (1996, pp. 20–25).

thinly pleased or displeased about something is to be pleased or displeased without feeling pleased or displeased. One can take displeasure in something—one can be displeased about something—without feeling any (sensory) pain; or one can be pleased about something without feeling any (sensory) pleasure. To be thickly pleased about something is, first, to be thinly pleased about it and, second, to feel pleased at it. (Similar things are true about thick displeasure.) When one is thickly pleased about something, one takes attitudinal pleasure in the relevant fact and feels pleasure. Zimmerman proposes that thick pleasure (feeling pleased about something) involves something like elation or euphoria, while thick displeasure involves something like depression or dysphoria. So feeling pleasure at something—thick pleasure—is being pleased at it plus feeling (something like) elation, and feeling displeasure at something is being displeased at it plus feeling (something like) dysphoria (Zimmerman 2001, p. 197). Zimmerman further proposes that it is possible to feel elated or depressed without being pleased or displeased about anything (2001, p. 196; 2007, p. 427). In addition, he claims that "when it is said that there is value in pleasure or disvalue in displeasure, it is surely *feeling* pleased or displeased that is usually intended. It is chiefly the euphoric aspect of (thick) pleasure that makes such pleasure pleasant, and the dysphoric aspect of (thick) displeasure that makes such displeasure unpleasant. There would seem to be no great personal benefit to be derived from being pleased when this involves no euphoria, no great personal cost in being displeased when this involves no dysphoria" (2007, p. 427).

There is a puzzle, though, with this last suggestion. I grant that certain affective states or feelings can be of personal benefit and other feelings can be of personal cost. That is clearly right. But if certain feelings can be of personal cost or benefit, can certain cognitive states (unaccompanied by affections) not be of personal benefit or cost, too? So, for instance, why can being thinly pleased about something not be of some personal benefit and being thinly displeased about something not be of personal cost? We might put the point in this way: just as it is good to feel elated, so it is good to be thinly pleased at, or about, or in something. There seems to be no reason to privilege affective states in the relevant fashion but exclude cognitive states from being privileged in this same fashion.

Now suppose you are thinly pleased about something. From the fact that you would not *feel* any different if you were not pleased about that thing, it does not follow that being thinly pleased about something is welfare-irrelevant.

Let us try again. When is or what makes some factor welfare-irrelevant but perhaps world-relevant? Maybe internal indiscernibility holds the key to distinguish between welfare-relevant and world-relevant factors. Introducing the notion of *affect** should help (the asterisk simply signals a technical sense of "affect"). Suppose we want to know whether truth is welfare-relevant; we want to know whether the object in which one takes intrinsic attitudinal pleasure's being true is welfare-relevant. We compare the segment of an agent's, *S*'s, life in which *S* takes pleasure in object *O* and in which *O* is false with an otherwise type-identical or near type-identical segment in which *S* takes pleasure in *O* but in which *O* is true. If the two segments are internally indiscernible, then the episode of intrinsic pleasure's lacking the factor in

question—in this instance, the object of the episode's being false—does not affect* *S*. It may now be put to us that

Proposal P: If an episode of *S*'s intrinsic attitudinal pleasure's lacking some factor, *f*, does not affect* *S*, then *f* is not welfare-relevant.

Here is an illustrative example. Unconscious beliefs may influence one's behavior or one's emotional life. But within a temporal duration, they might have no influence whatsoever on their agent. Assume that agent *S* has an unconscious belief of this sort. If we compare the segment of *S*'s life in which *S* has this belief with an otherwise type-identical (or near type-identical) segment in which *S* lacks this belief, the two segments are internally indiscernible. At any time within the temporal duration, lacking (or having) the belief does not affect* *S*.

Many attitudinal pleasures are *occurrent*. When a person takes such intrinsic attitudinal pleasure in some state of affairs at some time, at that very time, he is thinking about, and taking pleasure in that state of affairs. As beliefs can be unconscious, and beliefs are propositional attitudes, it appears just as plausible that many intrinsic attitudinal pleasures may not be occurrent but may in fact be unconscious. When a person takes such pleasure in something at a time, she is then taking pleasure in that thing but is unaware or does not believe at that time that she is then taking pleasure in that thing. *S's taking intrinsic pleasure in p* seems to entail that *S believes that p*.[4] If *S*'s taking pleasure in some object at some time is occurrent, then at the time *S* is taking pleasure in that object, *S* has the relevant belief; and presumably, this belief is occurrent. If *S*'s pleasure is unconscious, then *S* has the germane belief concerning the object of her pleasure and the belief may well be unconscious. Consider an episode of intrinsic attitudinal pleasure divorced from sensory pleasure. Suppose this episode is unconscious: *S* is unaware that *S* is taking pleasure in some object. If we compare the segment of *S*'s life in which *S* has this unconscious pleasure with an otherwise type-identical (or near type-identical) segment in which *S* lacks this pleasure, the two segments are internally indiscernible. Proposal P yields the result that *S*'s taking (thin) pleasure in something is welfare-irrelevant. But this is false, given attitudinal hedonism. If you take pleasure in various objects during a period of time, and all these episodes of pleasure are unconscious, I see no reason to deny that, though unconscious, these episodes contribute to your welfare level during the stretch of time. Hence, Proposal P ought to be rejected.

In sum, some people may be inclined to endorse the view that factors such as deception or unfreedom are welfare-irrelevant but world-relevant. It is difficult to discern what grounds this claim. Hence, I am far from convinced that the first objection against apt freedom-sensitive versions of attitudinal hedonism is anywhere close to knock-down.

---

[4] Lemos (1994, p. 78) disagrees; he thinks that in some cases one can take pleasure in some state of affairs without believing that it obtains.

3.6.2 Freedom and Manipulation

A second objection is that, even if one were to agree that Neo's life in Matrix 2 is not good in itself for Neo, this would be not simply because his intrinsic attitudinal pleasures are unfree but because they are manipulated by evil people. Again, however, this objection is not conclusive. Manipulation by evil people is not really relevant to the judgment that Neo's life (in Matrix-2) is not intrinsically good for Neo. The judgment would surely remain unaffected if, as a result of a series of incredible chance occurrences, Neo's embryo found itself housed in the slimy womb of the Matrix and electronic influences caused the individual into which this embryo develops to take intrinsic attitudinal pleasure in true objects. If we were to adjust hedonism to accommodate the judgment that Neo's life in the Matrix would not be good for Neo, how else could we do so but to adjust the values of episodes of intrinsic attitudinal pleasure according to which these episodes are free or not?

3.6.3 Freedom and Unbidden Pleasures

A third objection is that even if our world is replete with free and responsible *action*, the vast majority of attitudinal pleasures arise unbidden, and, hence, unfreely (unless merely indirectly and unintentionally freely). How can this diminish the value of such pleasures, let alone render them worthless (as the robust freedom-sensitive theory implies)?

As a first, initial comment, we have said that it is plausible that whether the object in which intrinsic attitudinal pleasure is taken is true can affect the value of the pleasure. This is so even if the pleasure is unbidden. But then why should one be so reluctant to grant that freedom can affect the value of an unbidden episode of pleasure? As a second, initial comment, it is not clear that the *vast majority* of our intrinsic attitudinal pleasures are unbidden. After all, we learn to take pleasure (and displeasure) in many things. One has to learn to take pleasure in appreciating certain paintings, in listening to various genres of music, or in savoring certain foods. With training, encouragement, or experimentation, we can stop taking intrinsic pleasure in things in which we formerly took pleasure, and we can begin to take intrinsic pleasure in things in which we formerly did not take pleasure or in which we took intrinsic displeasure. Scenarios such as the one Mele describes in which Tom takes intrinsic pleasure in Betty's failure but in which it seems that Tom has control over whether or not he takes such pleasure seem to be fairly common in everyday life.

Consider an intention that is passively (or nonactionally) acquired; it is not acquired as a result of *any* prior action such as making a decision. Sid's intention to sip the freshly brewed cup of coffee may be an intention of this sort as may be his intention to answer the phone when it rings. The intention may still be free although it could be acquired, at a relevant time, in a way in which its freedom is undermined. The freedom of such an intention is a function of the causal process that gives rise to it. Mele, for example, suggests that its nonactional freedom will be its freedom

from certain influences that would undermine the freedom of the subsequent action that it produces (if it does produce one), freedom from such things as compulsion, manipulation, and insanity (1995, p. 225).

It seems that many unbidden pleasures are relevantly like passively acquired intentions that are free and not pertinently like passively acquired intentions that are unfree: the causal processes that issue in the wholly unintentionally acquired mental states need not subvert their freedom. Still, one may not be satisfied with this first-stab response. It seems that many passively acquired mental states, just like many passively acquired intentions, will be precisely, and entirely, that—*passively* acquired; that is, not acquired by way of some (mental) *action* over which the agent has direct control. If entirely passively acquired, though, why would this enhance the value of an episode of intrinsic attitudinal pleasure? The answer that I favor is that, although unbidden but free in the manner explained, the agent is still implicated and is not merely on the sidelines, in the acquisition of the mental state. This requires explanation.

Previously, we said that *S's taking intrinsic pleasure in p* seems to entail that *S believes that p*. As I explain later, there is a difference between desires and beliefs that are "truly one's own" or "authentic" and those that are "inauthentic." Elsewhere, I have defended the view that in addition to having epistemic and freedom requirements, moral responsibility also has an authenticity requirement. Roughly, the requirement says that if a person is morally responsible for an action, the action causally derives from springs of action, such as beliefs, desires, values, and so forth that are authentic (see, e.g., Haji 1998, Haji and Cuypers 2004, 2008). Various forms of manipulation motivate this requirement. We can conceive of situations in which beliefs and desires are implanted into an individual without the individual's consent or knowledge. An action expresses a pro-attitude if that pro-attitude plays a nondeviant causal role in the production of that action. The individual's actions that express the implanted elements may be ones in the performance of which the individual exercises freedom-level control. Yet, if the case is carefully crafted, it reveals that it is highly implausible that the individual is morally *responsible* for such actions. This is because their relevant causal antecedents are not the "agent's own." (A case of this sort will be discussed in more detail in Chapter 5.)

A proponent of an authenticity requirement inherits the burden of advancing and defending an account of authenticity. When, for example, is a belief or desire authentic or inauthentic? Again, I will address elements of this issue later (in Section 5.1). For now, it is enough to note that if a belief or desire is acquired by a normal, healthy adult person as a result of a process that completely bypasses her capacities of deliberative control, then the belief or desire is inauthentic. In such instances, we can say that the person is not properly implicated in the acquisition of the belief or desire; the person's agency is, in this manner, bypassed in the acquisition. I propose that it is a good thing that in acquiring pro-attitudes, beliefs, deliberative principles, etc., one's agency not be bypassed.

Suppose some nefarious neurosurgeons make Cleo take intrinsic pleasure in various objects. If the state of affairs*, a person's taking pleasure in something, p*, entails *the person's having a belief that p*, the neurosurgeons will be successful only if they

implant in Cleo the relevant beliefs. These beliefs will be inauthentic; Cleo's agency is on the sidelines in their acquisition. If, to the contrary, Cleo takes intrinsic pleasure in these objects on "her own," the relevant beliefs are authentic (or at least assume, barring any extraordinary circumstances, that they are so). In this event, even if Cleo's intrinsic attitudinal pleasures are all unbidden, they are free in that they are derivatively authentic: their associated beliefs are all authentic. Since it is a good thing that one's actional elements are authentic, consequently it is not implausible to suppose that the freedom of even intrinsic pleasures that are unbidden enhances the value of these pleasures. There seems, then, to be no bar against holding that many of our unbidden pleasures, just like many of our passively acquired intentions (or beliefs), are free.

Suppose manipulators see to it that Neo takes intrinsic pleasure in various objects that are true by tinkering with Neo's attitudinal-pleasure-acquiring faculties. I am drawn to the view that Neo's life, though filled with unbidden episodes of attitudinal pleasure, is not so intrinsically good for Neo because these attitudes are unfree. If we grant that Neo's life is not so good in itself for Neo, we then have recourse to the following argument. Neo* is one of Neo's counterparts whose life is both internally indiscernible from Neo's and not a make-believe one. Although Neo*'s life is highly intrinsically good for Neo*, Neo's life is not so intrinsically good for Neo. But all of Neo's and Neo*'s intrinsic attitudinal pleasures are unbidden (or so we assume). So it is not this factor—that the pleasures are unbidden—that accounts for the difference in the intrinsic value of their lives. The only explanatorily relevant difference seems to be that Neo's unbidden intrinsic pleasures are not free, whereas those of Neo*'s are free.

3.6.4 Unfree Pleasures and Value

Suppose we grant that free intrinsic attitudinal pleasures are better than unfree ones. This does not imply that unfree intrinsic pleasures are not good to any extent. Hence, it is erroneous to judge that Neo's life in Matrix-2 is not intrinsically good to any extent for Neo. If cogent, this is an objection against *robust freedom-sensitive intrinsic attitudinal hedonism* that implies that unfree intrinsic pleasures and unfree intrinsic displeasures are intrinsically worthless, and against *pain-adjusted freedom-sensitive intrinsic attitudinal hedonism* that implies that unfree intrinsic pleasures (but not intrinsic displeasures) are intrinsically worthless. Even if sound, the objection leaves intact the symmetric versions of freedom-sensitive attitudinal hedonism. According to these versions, unfree intrinsic pleasures (and unfree intrinsic displeasures) can affect the degree to which an agent's life is valuable in itself for that agent (but such pleasures or displeasures are not intrinsically worthless).

Whether this fourth objection against, for example, the robust theory *is* sound turns principally on whether Neo's life in Matrix-2 is intrinsically good (to some degree) for Neo. It is challenging, to say the least, to argue categorically for the conclusion that Neo's life is not intrinsically good. But some may think that it is not implausible to maintain that this is so. Hedonic act utilitarianism is, roughly, the normative theory that an act is morally obligatory if and only if its total outcome

contains higher net pleasure than the net pleasure contained in the total outcome of any alternative. As some may see it, the disagreement between those who hold that Neo's life (in Matrix-2) is good to some extent and those who hold that Neo's life is intrinsically worthless is somewhat analogous to the disagreement between those who maintain that hedonic act utilitarianism is defective because, under certain circumstances, it sanctions framing an innocent person for a crime that he did not commit and those who staunchly hold fast to the normative theory despite what others see as this undermining result. My view is that the argument from injustice is a telling objection against hedonic act utilitarianism. If this objection does not cast doubt on the theory, it is difficult to see what sort of moral objection would do so. But the objection gives us no reason to reject, wholesale, the utilitarian insight that we ought to do the best we can. Rather, as Feldman (1997, pp. 154–174) has proposed, one should acknowledge the force of the objection and preserve this utilitarian insight by adjusting utility for justice.

In like fashion, if unfree episodes of intrinsic attitudinal pleasures and intrinsic displeasures *did* have intrinsic value, Neo's life in Matrix-2 would be good in itself for Neo. Again, this is because his make-believe life contains numerous episodes of intrinsic attitudinal pleasure and no episodes of intrinsic displeasure. But some may find it hard to swallow the result that Neo's life in his slimy womb is good in itself for Neo just as it is hard to swallow that it is morally obligatory to frame the innocent person because doing so maximizes utility. Perhaps it might be suggested that unless one has special reasons to think otherwise, or unless one is desperate to hold on to a version of attitudinal hedonism that implies otherwise, it is difficult to imagine why anyone would want to affirm that Neo's life as described would be intrinsically good for Neo. On any standard, this is a terrible life for the person who lives it. It may be well worth emphasizing that one need not be confusing world-relevant factors with welfare-relevant ones in arriving at this verdict. Nor need one be confusing Neo's being happy with Neo's welfare level. Still, it may be claimed that the objection from unfreedom gives us little reason to reject outright the hedonist's insight that intrinsic attitudinal pleasures and intrinsic displeasures contribute to the intrinsic value of lives for persons. Finally, we might be invited to acknowledge the force of this objection and preserve this insight by adjusting the values of episodes of intrinsic attitudinal pleasures and displeasures according to whether such episodes are free.

But I think there is a problem with this view. Suppose we are hedonists of sorts and we acknowledge that Neo in Matrix-2 is happy. If he is happy, surely his being so is at least partly due to the fact that he enjoys intrinsic attitudinal pleasure. But it is unclear why his enjoying attitudinal pleasure would contribute to his happiness if his (unfree) episodes of intrinsic pleasure were all entirely devoid of intrinsic value. Thus, I am inclined to accept this objection against the pertinent forms of freedom-sensitive attitudinal hedonism. But as we remarked earlier, this objection does not impugn the symmetric versions of this form of hedonism.

Assuming that the robust theory and the pain-adjusted one are not in the running, this leaves the asymmetric theory and the two symmetric versions. On the asymmetric theory, freedom enhances the intrinsic value of episodes of intrinsic

attitudinal pleasure but has no effect on the intrinsic value of episodes of intrinsic displeasure. I find it prima facie implausible that freedom may have such an effect on intrinsic pleasures but no impact whatsoever on intrinsic displeasures. So I am inclined to settle for one of the symmetric theories. I have claimed that I prefer the symmetric version according to which free intrinsic pleasures are better than otherwise similar pleasures that are unfree and that free intrinsic displeasures are not as bad as otherwise similar displeasures that are not free.

Chapter 4
Pleasure, Desert, and Welfare

4.1 Pleasure and Desert

So far, I have been developing the view that freedom affects the intrinsic value of life atoms; it affects the value of episodes of intrinsic pleasure and displeasure, the sum of these values being a measure of how good in itself a life is for the person who lives that life. With a freedom-sensitive attitudinal hedonism as the relevant axiology, on the view that I favor, free intrinsic attitudinal pleasures are better than otherwise similar unfree pleasures, and free intrinsic displeasures are not as bad as otherwise similar unfree displeasures. In this chapter, I want to lay the groundwork for a slightly different route to the more cautious conclusion that freedom may have a bearing on welfare. In roughly hewn strokes, the idea is this: deserved pleasures are better than otherwise similar pleasures that are not deserved. Deserved displeasures are not as bad as otherwise similar displeasures that are not deserved. Freedom affects desert; so freedom affects the intrinsic value of pleasures and displeasures. Suppose the value of episodes of pleasures and displeasures are adjusted to reflect the extent to which they are deserved, and suppose that aptly specified atoms of attitudinal pleasure and displeasure whose values have been so adjusted are indeed *life* atoms—they are the fundamental "determinants" of *welfare*; then, once again, we will have shown that freedom can affect welfare.

Starting with some observations about desert, there are many different desert bases; there are, that is, many different factors that affect the extent to which a given person deserves a certain pleasure or displeasure: excessive or deficient past receipt, moral worthiness or virtue, legitimate claims, established character, etc. (see, e.g., Rescher 1966, pp. 73–83). Consider virtue (or viciousness). Among other things, a virtuous person is a person who habitually acts "from" virtue; similarly, a vicious person habitually acts "out of" vice. You may be deserving of pleasure if your acts exemplify virtue—you perform virtuous deeds; you may be deserving of displeasure if your acts express vice—you perform vicious deeds. Then it seems that in this fashion, virtuous and vicious deeds (actions or intentional omissions) are "desert bases." One's actions, though, need not express virtue for it to be true that, because of these deeds, one deserves pleasure. You may give alms to the poor in the belief that, in so doing, you do what is morally obligatory. You may well be deserving of

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

moral praise for your deed even if this act does not spring from virtue; equally, you may be deserving of pleasure.

A person can deserve pleasure for many reasons: she may have performed many morally good deeds; she is innocent and maybe innocent people deserve pleasure in virtue of their innocence; she has been deprived of food and she deserves the pleasures of a good meal, and so on. Analogously, a person may deserve displeasure for many different reasons. Needless to say, a person may, at a time, deserve pleasure for some reasons and may also, at that time, deserve displeasure for other reasons or may receive pleasure or displeasure that is undeserved. "Undeserved," it should be cautioned, masks an ambiguity that is worth exposing. It may be taken to mean the same as either "not deserved" or "deserved not." Usually, it is taken to mean the latter. We may bring out this distinction by introducing another set of distinctions concerning desert bases. A person has a *negative desert base* if the person deserves displeasure; a person has a *positive desert base* if she deserves pleasure; a person has a *neutral desert base* if she neither deserves pleasure nor deserves displeasure. Suppose a person with neutral desert base receives some pleasure. It would be correct to say that she does not deserve to receive the pleasure that she receives; there would be no good reason, though, to say that she *deserves* not to receive the pleasure she receives.

Now consider these axiological principles that govern the values of episodes of attitudinal pleasure (and displeasure) that are deserved or not deserved:

> AXP1: Positive desert base enhances the intrinsic value of an episode of attitudinal pleasure (Feldman 1997, p. 163). If one receives the pleasure that one deserves, the value of that episode of attitudinal pleasure is enhanced.

We record a few things about AXP1. First, AXP1 is relevantly analogous to the widely accepted principle that pleasure in the good is intrinsically good. If someone takes pleasure in the good, and this person deserves this pleasure because of past good deeds, then the intrinsic goodness of such an episode seems to be enhanced (the pleasure is made better) by virtue of his getting what he deserves (see, e.g., Moore [1903] 1962, p. 224; Smart 1973, p. 24; Chisholsm 1986, p. 63; Lemos 1994, p. 74; and Zimmerman 2001, p. 220).

Second, Feldman's view concerning the content of this principle can be better expressed in this way:

> AXP1-F: If one receives the pleasure that one deserves, the value of this episode of attitudinal pleasure *for the world* is enhanced.

It is important to appreciate the import of the italicized phrase in AXP1. Feldman proposes that if a person fully deserves a certain good and receives that good, then the value of the *world* is substantially increased. The value for the world of his deserved receipt, owing to the perfect match between receipt level and desert level, is greater than the value received by the person, taken just by itself. I formulate the principle in the fashion that I do because I am interested in whether it is true that if desert has an effect on *well-being*, freedom does so as well via freedom's influence on desert.

Third, this axiological principle does not take into account whether the pleasure is taken in an object that deserves to have pleasure taken in it. One might, for example, insist that deserved pleasure in someone's pain or in someone's undeserved pain is intrinsically bad.[1] (The other four axiological principles to follow also ignore object-worthiness.) Such adjustments for object-worthiness will not affect one of the primary points for which I wish to argue: determinism can have a detrimental impact on the value of *worlds*.

Moving on to the next principle,

> AXP2: Negative desert base mitigates the intrinsic value of an episode of displeasure (Feldman 1997, pp. 164–165). If one receives the attitudinal displeasure that one deserves, the value of that episode of displeasure is mitigated.

One's receiving displeasure is (other things being equal) intrinsically bad. But, again, it seems highly plausible that if one receives some displeasure that one deserves, this is not *so* bad; the value of the episode of displeasure is mitigated (the displeasure is made less bad).

> AXP3: Neutral desert base neither enhances nor mitigates the value of pleasure or displeasure (Feldman 1997, pp. 166; 168–169). The intrinsic value of an episode of pleasure (or displeasure) of this sort. is directly proportional to the amount of pleasure (or displeasure) it contains.

Suppose you do not deserve any pleasure and you do not deserve any displeasure but you receive some pleasure and displeasure. It appears that the values of the episodes of gratuitous pleasure (and displeasure) you receive mirror the amount of pleasure or displeasure in those episodes.

> AXP4: Negative desert mitigates the intrinsic goodness of pleasure. If someone deserves displeasure but gets pleasure instead, the value of that pleasure is mitigated (the pleasure is made less good) (Feldman 1997, pp. 164–165).

Finally,

> AXP5: Positive desert aggravates the intrinsic badness of displeasure. If someone deserves pleasure but gets displeasure instead, the value of that displeasure is aggravated (the displeasure is made yet worse) (Feldman 1997, pp. 166–167).

Let the values of the episodes of intrinsic pleasures and intrinsic displeasures that a subject experiences be adjusted for desert in accordance with principles AXP1–AXP5. Feldman dubs this new measure of value "subject's desert-adjusted intrinsic value." He claims that *subject's desert-adjusted intrinsic attitudinal hedonism* (SDAIAH) is the view that the value of *a world* (but not of a life) is the sum of the subject's desert-adjusted values of the intrinsic attitudinal pleasures enjoyed and intrinsic displeasures suffered in that world. The theory runs as follows:

> Subject's desert-adjusted intrinsic attitudinal hedonism (SDAIAH)
>
> i. Every episode of intrinsic attitudinal pleasure is intrinsically good; every episode of intrinsic displeasure is intrinsically bad.

[1] For some thoughts on deserved pleasure in someone's pain, see McLeod (2006).

ii. The subject's desert-adjusted intrinsic value of an episode of intrinsic attitudinal pleasure is equal to the amount of pleasure contained in that episode adjusted for subject's desert; the subject's desert-adjusted intrinsic value of an episode of intrinsic displeasure is equal to –(the amount of displeasure contained in that episode adjusted for the subject's desert).
iii. The intrinsic value of a world is entirely determined by the subject's desert-adjusted intrinsic values of the episodes of intrinsic attitudinal pleasure and displeasure contained in that world, in such a way that one world is intrinsically better than another if and only if the net amount of intrinsic attitudinal pleasure adjusted for subject's desert in the one is greater than the net amount of that sort of pleasure in the other (Feldman 2004, p. 195).[2]

On this theory when we say that a pleasure is made better if (for instance) the person who receives it deserves to receive it, we are speaking somewhat loosely. We should not be taken to be claiming that pleasures have variable intrinsic values—values that can be increased or decreased depending upon whether subjects who receive them deserve or do not deserve to receive them. Rather, when we say things of this sort, we are strictly expressing some fact about the basics on the desert-adjusted theory. A basic intrinsic value state on this axiology would be a state of affairs of this form: *S takes intrinsic attitudinal pleasure (displeasure) of intensity n1 and duration m1 at time t in state of affairs P when S deserves to degree r1 to be taking that pleasure (or displeasure).*

4.2 Why Be Drawn to Subject's Desert-Adjusted Intrinsic Attitudinal Hedonism?

Feldman underscores the point that SDAIAH offers an evaluation of *worlds* unlike, for example, the simple theory or an attitudinal hedonism that adjusts the values of episodes of intrinsic attitudinal pleasures and displeasures for object-worthiness. This latter sort of hedonism, he says, may offer an evaluation of *lives*. According to Feldman, then, world atoms differ from life atoms; "atomism"—the view that there is a uniform set of atoms for the assessment of worlds, lives, the total consequences of actions, and so forth—is false. One might accept the falsity of atomism and yet wonder why it may not also be true that whether a subject deserves (or fails to deserve) a pleasure or displeasure contributes to the value of that person's *life*. It is this issue that I wish to explore. But before I do so, we should say something about at least one strong motivation behind SDAIAH. (Later, in Section 7.7.2, we will discuss another incentive to endorse SDAIAH, one having to do with avoiding the so-called "repugnant conclusion.")

---

[2] Assume that the number of basic intrinsic value states true at a world is finite. If the number were infinite, clause (iii)'s summative principle would need recourse to a more complicated mathematics.

4.2.1 Ross's Two-Worlds Objection

Simple intrinsic attitudinal hedonism (SIAH) is a life- and not a world-ranking axiology. In Chapter 2, we considered a number of objections against this simple theory. Each was an objection to what some have taken to be unacceptable implications of the theory's assessment of various lives. There is a variant of the simple theory that is a world-ranking axiology. To formulate this theory—SIAH-W—we simply amend the third clause of SIAH as follows:

iii. The intrinsic value of a world is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and displeasure contained in the world, in such a way that one world is intrinsically better than another if and only if the net amount of intrinsic attitudinal pleasure in the one is greater than the net amount of intrinsic attitudinal pleasure in the other.

W.D. Ross raises a potent objection against world-ranking forms of hedonism (such as SIAH-W) that involves a comparison of two worlds, a just one, W_j, and an unjust one, W_k (Ross 1930, p. 138). It is stipulated that each of these worlds contains exactly the same net amount of pleasure. In some versions of the "two-worlds" objection, the pleasure is sensory; in other versions, the pleasure is attitudinal. Assume, for present concerns, that the pleasure (and pain) are attitudinal and the target of Ross' objection is SIAH-W. In W_j, all the pleasures are appropriately distributed to those who deserve the pleasures and all the displeasures are fittingly distributed to those who deserve the displeasures. In W_k, the distribution of pleasures and displeasures is reversed: the people who deserve displeasure get pleasure, while those who deserve pleasure get displeasure. Given certain assumptions, SIAH-W sustains the verdict that each person in W_k lives a life that is highly good in itself for that person. Still, W_k is not a good *world*. More cautiously, the less fitting allocation of pleasures and displeasures in W_k makes it worse in itself than W_j.

In an important passage, Franz Brentano draws a distinction that bears fairly directly on the two worlds objection:

> We may distinguish between realizing what is good in general, a good in the whole world-order, and realizing a good in a particular individual. If at the Last Judgement a greater amount of bliss were given to a person who actually deserved it less, then he would have a greater amount of good than he otherwise would have, but the good in the universe, considered as a whole, would be less. (1969, p. 149)

Suppose some person receives a primary intrinsic good, in Brentano's case, an episode of bliss. Brentano distinguishes between the intrinsic value in that episode for the recipient and the intrinsic value in that episode for the world as a result of its recipient's receiving the good. With the latter evaluation, we consider both the amount of the good that the recipient receives and the extent to which the recipient deserves the good. In the case that Brentano explicitly discusses, someone enjoys a good that he does not deserve. Although the person might be better off—his welfare level might be enhanced—as a result of his enjoying the good, his receiving the

good does not make the world better off because there is a mismatch between what the person deserves to receive and what the person receives. We may think of the mismatch as an injustice. Then we can summarize Brentano's point in this way: the injustice, owing to there being a mismatch between receipt level and desert level, makes the person but not the world better off. Axiological principles AXP1–AXP5 can be thought of as telling us how the value for the world is affected in other cases in which someone receives a primary intrinsic good. Each of these principles is sensitive to both receipt level and desert level.

We may now better appreciate why a world-ranking attitudinal hedonism that is sensitive to how pleasures and displeasures are distributed—a hedonism that, in effect, adjusts the values of episodes of intrinsic pleasures and intrinsic displeasures in accordance with whether these episodes are deserved or not deserved—holds the promise of yielding the result that the just world (W_j) is better than the unjust world (W_k). SDAIAH is just this sort of hedonism. It looks to how receipt of episodes of intrinsic attitudinal pleasures and intrinsic displeasures, when conjoined with facts about whether the episodes are deserved or not deserved by their recipients, affects worldly value. Consonant with Ross' estimation, *SDAIAH* (unlike *SIAH-W*) rates the just world as intrinsically superior to the unjust world.

4.3 On the Value of Worlds and Lives

Do atoms of pleasure and displeasure, whose value is modulated to reflect the extent to which these pleasures are deserved or not deserved, have a bearing not just on the value of worlds but also on the value of lives? Reconsider the sort of hedonism—"Object-Worthy Hedonism"—that adjusts the value of an episode of intrinsic attitudinal pleasure (or intrinsic displeasure) in accordance with whether the object of that pleasure (or displeasure) deserves to be enjoyed. Feldman, as we have documented, claims that some may think that object-worthiness has an impact on the intrinsic value of lives. Suppose Spike and his twin, Pike, have no special aesthetic training. Spike takes great intrinsic pleasure in viewing the Mona Lisa; the object of his pleasure is genuinely beautiful and it deserves to be appreciated. Pike takes similar pleasure in a copy of the Mona Lisa that he believes is the masterpiece. The counterfeit is not "object-worthy"; it is not an object that deserves to be an object of pleasure. Assume that Spike would have displayed the same sort of enjoyment that he displayed in gazing at the real thing had he, instead, been presented with the fake. Assume, further, that corresponding things are true about his twin. *Object-worthy hedonism* implies that the value of Spike's pertinent episode of intrinsic pleasure is enhanced in virtue of his taking pleasure in an object that is object-worthy, and it implies that the value of Pike's pertinent episode of intrinsic attitudinal pleasure is not so enhanced. This is so even if the relevant life segments of these two individuals are internally indiscernible—the segments containing the "aesthetic pleasures" of each of the twins may "feel" exactly the same "from the inside."

On route to inquiring whether subject's desert affects the value of lives, we first ask whether indiscernibleness of this sort calls into question *object-worthy hedo-*

nism. The concern is that since object-worthiness need have no influence whatsoever on how one's life "feels" from the "inside," object-worthiness is not germane to well-being (though it may be germane to the value of worlds). I am inclined to think that this concern is not on target. *Why*, after all, should it be the case that whether some factor, such as object-worthiness, makes a difference to how one's life "feels" from the "inside" be the test of whether this factor is pertinent to an evaluation of lives rather than to an evaluation of worlds?

Perhaps, one might claim that we should distinguish between personal value and impersonal (or ethical) value. Something is personally good for some person if and only if it is good in terms of the welfare or well-being of that person. Ethical value is one type of nonderivative value. Michael Zimmerman explains

> When Ross and Feldman. . .say that the world in which the virtuous prosper and the vicious suffer is better than the world in which the reverse is true, they are (I believe) looking at matters from an ethical standpoint. It is *ethically* fitting that personal goods and evils be distributed as they are in the first world, *ethically* unfitting that they be distributed as they are in the second world; hence, an ethically sensitive person would, *ceteris paribus*, prefer the first world to the second (2007, p. 429).

Regarding the sort of adjustment to the values of episodes of intrinsic attitudinal pleasure and intrinsic displeasure that Object-Worthy Hedonism recommends, Zimmerman questions why we should bother with the adjustments in the first place, since Moore's objection from base pleasures seems, often, to stem from *ethical* concerns:

> [S]omeone who disapproves of a life full of "low" pleasures is likely, I think, to be questioning the *ethical* value of such a life, rather than the claim that such a life is good in terms of *personal* welfare [note omitted]. But if this is so, then the proper response to the objection is not to absorb it by adjusting one's theory but to reject it outright as misdirected, since it concerns ethical value and not. . .personal value. . . (2007, p. 431).

In Zimmerman's assessment, Moore's objection gives us reason to believe that a life replete with intrinsic attitudinal pleasures taken in unworthy objects may well be low in ethical value but not in personal value, and it is the latter and not the former that is pertinent to well-being.

While I find Zimmerman's reflections highly suggestive, I do not think that they provide conclusive grounds to unseat *object-worthy hedonism*. Grant the distinction between personal value and ethical value. At issue, really, is whether object-worthiness may bear on *personal value* even if it is allowed that object-worthiness may bear on ethical value. One need not be confusing the two sorts of value in recommending that the extent to which an object deserves to have pleasure taken in it may have definite impact on personal value.

In sum, Moore advances an objection—the objection from base pleasures—that may reasonably be regarded as an objection against the life-ranking simple theory (SIAH). On one response, Moore's objection misfires; it conflates personal value and ethical value. I prefer to be more charitable: the Moorean objection resurfaces even if one does not confuse these two sorts of value. On Feldman's second response (perhaps) or on an alternative response, the proponent of *object-worthy hedonism*

construes Moore's objection as one against an axiological theory about personal value. An implication of this second response is that there are "subject-independent factors"—roughly, factors that need not in any way influence what the life of a person "feels" to that person from the "inside" and that affect well-being. Object-worthiness is one such factor. Others have suggested that truth is another such factor.

We may now finally revert to our initial riddle: why should it not be the case that if a subject deserves (or fails to deserve) an intrinsic pleasure (or an intrinsic displeasure), this factor contributes to that person's well-being—it contributes to personal value—even if it is acknowledged that it contributes to the intrinsic value of worlds? Feldman submits that the evaluation of worlds and the evaluation of lives make use of different considerations:

> If we want to know how well a person's life is going for him, we want to know the net extent to which he is enjoying things. On the other hand, if we want to know how well thing are going in a world, we want to know something about the extent to which people are enjoying good things and suffering bad things, taking account of the extent to which those people deserve to be enjoying the good things and suffering the bad ones (2004, p. 195).

I find the view that different considerations pertain to the evaluation of lives and of worlds attractive. Indeed, there *are* prima facie powerful reasons to believe that desert (or justice) may be welfare-irrelevant but world-relevant. Suppose Alvin commits a crime, say a murder, he deserves the displeasure of long-term confinement and he does in fact suffer such confinement. He experiences displeasure in accord with what he deserves. But even if we accept that the actual world is intrinsically better ceteris paribus, owing to this instantiation of justice, than a pertinent counterfactual world in which Alvin commits the murder, is not caught and thus does not suffer the relevant displeasures; it is far from pellucid that Alvin's life in confinement is better for Alvin than his life is for him in the counterfactual world.[3] However, there is still the quandary that we touched upon previously: precisely what makes some factor welfare-irrelevant but world-relevant? I prefer to err on the side of caution. If we grant that a subject-independent factor, such as object-worthiness, may bear on the evaluation of lives, it is not transparent why we should deny that another subject-independent factor, such as subject's desert, may not also bear on the evaluation of lives (even having conceded that it may bear on the evaluation of worlds), given principles such as AXP1–AXP5 and no decisive view concerning the features a factor must satisfy to be welfare-irrelevant.

I register an assumption:

> *Assumption Value*: If object-worthiness (or truth) can have an impact on the value of episodes of intrinsic attitudinal pleasures and displeasures, and in virtue of this impact can affect well-being, then subject's desert can also have an impact on well-being by way of having an impact on the value of episodes of intrinsic attitudinal pleasures and displeasures.

I believe that this assumption is credible. If it is true, then, as I argue below, the freedom of episodes of intrinsic attitudinal pleasures affects the well-being

[3] I owe to this example to Al Mele whom I thank. For similar examples, see Feldman (2004, p. 196).

via considerations of desert. More specifically, the reasoning unfolds roughly in this way:

1. Deserved pleasures are better than otherwise similar pleasures that are not deserved and deserved displeasures are less bad than otherwise similar displeasures that are not deserved.
2. A pleasure (or displeasure) of a certain sort—roughly one that if deserved is deserved because of an action that one performs—is deserved only if it is free. These are "action-based" pleasures.
3. If (1) and (2) are true, then freedom can have an impact on the value of intrinsic pleasures and displeasures.
4. If freedom can have an impact on the value of these things, then given *assumption value*—specifically, given that subject's desert can have an impact on well-being by having an impact on the value of episodes of intrinsic pleasures and displeasures—freedom can have an impact on well-being. Thus, freedom can have an impact on well-being.

If it turns out that *assumption value* is false, this will not affect the reasoning for another conclusion to be argued for in Section 6.5: freedom affects the value of *worlds* because freedom has a bearing on desert of pleasures and displeasures.

4.4 The Freedom of Our Decisions

I previously proposed (in Section 3.1) that if episodes of pleasure and displeasure are free, they are indirectly free. These episodes are indirectly free only if they causally arise from decisions that are directly free. To support the view that an action-based pleasure (or displeasure) is deserved only if it is free, first we need to say something about when decisions are free. In other works, I have defended the view that a decision is free if and only if its agent has control in making it and he is autonomous with respect to it. The control and autonomy component each merits independent scrutiny.

As decisions are mental actions, the control condition can be parsed as follows. If a mental action is a decision, this action is free only if its agent has control in performing it. The control at issue is the control that moral responsibility demands. As we have already noted, competing accounts of in what this control consists have been proposed. Many libertarians say that the ability to do otherwise is required for moral responsibility; others claim that to be morally responsible, one must be appropriately sensitive to reasons; yet others defend the view that responsibility requires some sort of hierarchical control; we must have the ability to identify with our first-order desires via second-order volitions if we are to be responsible for behavior that causally issues from our first-order desires; and so on. Needless to say, this is not the place to assess these or other substantive accounts of control. That free decisions must meet *some* control condition is not controversial.

As far as autonomy is concerned, "autonomy" has at least two different nuances—one associated with *being self-governing* and the other with *ownership*. It is the latter overtone that is salient to the autonomy component of free decisions. I suspect that the former is, in the minds of many, more closely aligned either with the control component (where the stress is on "govern*ing*") or with *both* control and autonomy.

The autonomy component says that a decision is free only if it issues from antecedent springs of action, such as desires and beliefs, which are "authentic" or "truly the agent's own." Varieties of manipulation that undermine agency or moral responsibility bring out the pertinent contrast between, for instance, pro-attitudes such as desires that are authentic and those that are inauthentic. Globally manipulated agents are agents who, unaware of being finagled with, fall victim to manipulation that results in significant alteration of their psychological constitution. The "implanted" pro-attitudes are practically unsheddable. As Mele (1995, p. 172) explains, a pro-attitude is practically unsheddable for a person at a time if, given her psychological constitution at that time, ridding herself of that attitude is not a "psychologically genuine option" under any but extraordinary circumstances. In one of Mele's examples of such wholesale psychological manipulation, Ann and Beth are both philosophy professors but Ann is far more dedicated to the discipline. Wanting more production out of Beth and not scrupulous about how he gets it, the dean of the University procures the help of new-wave neurologists who "implant" in easy-going Beth Ann's hierarchy of values. The global induction results in Beth's being, in *relevant respects*, the psychological twin of Ann (Mele 1995, p. 145).[4] So, for instance, the induction leaves unscathed values, beliefs, desires, and so forth, which premanipulated Beth possessed and which can coexist more or less harmoniously with the implanted pro-attitudes. Such psychological tempering is thus consistent with preserving personal identity: premanipulated Beth is identical to her postmanipulated later self.[5] Should, though, the story generate qualms about personal identity, scale down the manipulation until one is confident that pre- and postmanipulated Beth are one and the same person.

---

[4] On global manipulation, see, for instance, Locke (1975), Kane (1996), Fischer and Ravizza (1998), esp. Chapters 7 and 8; Pereboom (2001), pp. 110–117; and Haji and Cuypers (2007, 2008).

[5] Walter Glannon (1998, especially Section IV) has argued that our practices of holding people morally and criminally responsible require only a low threshold of psychological connectedness and bodily continuity. In 1995, p. 175, n. 22, Mele suggests that in such transformation cases, the pre- and postsurgery agents may be strongly psychologically connected, in Parfit's (1984, p. 206) sense. They may be such that the number of direct psychological connections between them "is *at least half* the number that hold, over every day, in the lives of nearly every actual person." In addition, Mele argues that the presurgery agent (t-Beth) just before her transformation is much more similar, on the whole, to the postsurgery agent (t*-Beth) than she is to neonate Beth or toddler Beth. Still, t-Beth is the same person as the neonate and toddler Beths, in a familiar "personal identity" sense of "same person." So what is to prevent her from being the same person, in the same sense, as t*-Beth? On the supposition of personal identity, see also, Haji (2000a).

Regarding the first few engineered-in, practically unsheddable desires that move victimized Beth to action, we would want to say that these desires are not truly Beth's own. Many would not, for instance, agree that manipulated Beth is morally responsible for an action that expresses such desires even though she may exercise the *control or freedom* that moral responsibility requires in performing this action. Indeed, one alleged moral of global induction cases is that, in addition to, for instance, control and epistemic requirements, responsibility has ownership (or authenticity) requirements as well (see, e.g., Mele 1995, 2006; Haji and Cuypers 2008). I am proposing that there is a parallel authenticity requirement on well-being.

Elaborating, spin the Ann/Beth tale in such a fashion that it is clear, first, that manipulation-free Ann satisfies the favored freedom-relevant condition (whatever it may be), together with other conditions of responsibility, such as the epistemic one, in performing a certain type of action. Bracketing off certain complications concerning the compatibility of free action or responsibility with determinism (or indeterminism), we would judge that Ann is morally responsible for the deed at issue. Next, imagine that as a result of manipulation, Beth performs a type- or near-type identical action to the one that Ann performs. A highly plausible judgment is that manipulated Beth is *not* responsible for performing this action even though she satisfies whatever conditions on responsibility (save, of course, for the ownership or authenticity one) that Ann satisfies. Beth is too much a victim of her manipulators to be morally responsible for the pertinent deed.

Merely stipulating that Ann's (and so Beth's) relevant action issues indeterministically from antecedent springs of action should have no effect on one's verdict of nonresponsibility regarding manipulated Beth provided one judges that in deterministic scenarios, manipulated Beth is not responsible. Simply imagine that the manipulators ensure that there is a very high probability that Beth will perform the type- or near-type identical action that Ann does and a very low probability that Beth will refrain from doing anything at all. In the indeterministic version of the story, Beth performs the type- or near-type identical action that Ann does.[6]

In both determinstic and indeterministic versions, whereas manipulated Beth exercised no control in the processes that gave rise to her Ann-like constellation of pro-attitudes, beliefs, and so forth, Ann apparently exercised significant control in acquiring and maintaining these things. This sort of difference seems salient in informing the asymmetrical verdicts concerning responsibility that the scenarios elicit and (as we will see) plays a significant role in accounting for why manipulated Beth's desires, unlike Ann's, are inauthentic.

I concur that any proponent of the sort of view of free decisions that I have advanced inherits the burden of defending its authenticity component and of supplying an account of this component that illuminates the distinction between authentic and inauthentic actional springs. In the interests of not being too far detracted from

[6] For elaboration on this point, see, for instance, Haji (2000b), Mele (2006, pp. 138–144); and Pereboom (2002, pp. 477–488).

the issue of immediate concern—whether freedom can affect welfare as a result of affecting desert—I defer giving an account of authenticity to the next chapter. Here, I simply register that many free-will theorists contributing to the pertinent literature both acknowledge the distinction between authentic and inauthentic springs of action and endorse the view that free decisions have an authenticity component (see, e.g., Mele 1995, Fischer and Ravizza 1998). At least one welfare theorist, Sumner (1996), has proposed that authenticity is pertinent to welfare (some of Sumner's relevant views are examined in Chapter 12). In the next three sections, I proceed as follows. I first clarify the connection between the authenticity component of free decisions and the values of what a freedom-sensitive hedonism recognizes as life atoms. I then utilize the authenticity component, in tandem with axiological principles AXP1–AXP5 that link the values of episodes of pleasure (and displeasure) with whether these pleasures (and displeasures) are deserved or not, to provide additional motivation for the view that the freedom or unfreedom of a pleasure (or displeasure) can have a definite bearing on the intrinsic value of that pleasure (or displeasure).

4.5 Authentic Springs of Action and Value

I have proposed that provided that our lives contain some intrinsic attitudinal pleasure and some intrinsic displeasure, other things being equal, our well-being is maximized if our pleasures are free. To be more circumspect, other things being equal, the life of an agent that contains episodes of intrinsic attitudinal pleasure all of which are free is intrinsically better for that agent than an otherwise similar life in which the episodes of pleasure are unfree. If one favors the version of the symmetric theory that I do according to which free pleasures are better than otherwise similar unfree pleasures, and free displeasures are not as bad as otherwise similar unfree displeasures, then, again, if our lives contain some intrinsic pleasures and displeasures, we can arrive at a similar conclusion: other things being equal, a life in which all of one's pleasures and displeasures are free is better for one than an otherwise similar life in which the episodes of pleasure and displeasure are unfree. I have also submitted that the (indirect) freedom of a (free) pleasure traces to the direct freedom of a decision that gives rise to that pleasure; a pleasure is free only if it causally derives from a decision that is free. Next, I explained that a decision, in turn, is free only if it causally issues from authentic causal antecedents such as desires and beliefs; its agent is autonomous with respect to these proximal antecedents. Let us now in addition say that a mental action, such as a decision, is *derivatively authentic* if and only if it causally arises from authentic springs of action. Springs of action that are not authentic, as well as decisions that are not derivatively authentic, are inauthentic. Given these premises, we can, hence, conclude that

Free Pleasure: A pleasure is free only if it causally issues from authentic springs of action.

Corresponding things are true with displeasures. It is in this fashion that the atoms of value on a freedom-sensitive hedonism are essentially associated with authentic springs of action.

An objection to this sequence of reasoning targets the proposal that a decision is free only if it is (derivatively) authentic—that is, only if the decision causally arises from pertinent causal antecedents that are themselves authentic. For it seems that free decisions need not be so authentic. Many decisions may, for instance, ultimately derive from the uncritical acceptance of principles, values, or deliberative principles that have currency in one's culture or social milieu, and they may do so because a number of the pertinent pro-attitudes, such as the desire to act in conformity with traditional values, may be acquired during early childhood when the child lacks even minimal capacities of rational deliberation. Even so, it may be rejoined, subsequent decisions that nondeviantly arise from such pro-attitudes, in the absence of things like coercion and other factors that are obviously control-undermining, would be free. We can articulate the problem in a different way. I have proposed that a decision is free if and only if it is derivatively authentic and satisfies the correct condition of control (whatever this may be). How can we square this proposal with the view that free decisions need not be derivatively authentic because they may derive from relevant pro-attitudes whose acquisition bypasses the agent's (even perhaps minimal) capacities of reflective control?[7]

A full reply would draw on the substantive account of the authenticity of our springs of action. The broad contours, though, of an apt response to the objection are not difficult to discern if we are careful to differentiate two senses of "free" in which a decision may be free. In the strong sense pertinent to moral responsibility, a decision is strongly free if and only if it is derivatively authentic and satisfies the correct control condition. In a weak sense, a decision is weakly free if it satisfies the correct control condition. Globally manipulated Beth's posttransformation decisions are weakly but not strongly free. Reconsider the challenge that a decision may be free even though it derives from values, deliberative principles, and so forth, which have been acquired as a result of processes that bypass the agent's capacities of deliberative control. Regarding the decision in question, what is being envisaged, it appears, is a decision analogous to the ones that posttransformation Beth makes. We may grant that there is a sense in which such a decision is free: it is weakly free. But the decision, because derivatively inauthentic (it causally derives from springs that are inauthentic), is not strongly free; it is not one for which its agent is deserving of, say, moral praise or moral blame.

If free intrinsic attitudinal pleasures are better than corresponding unfree ones, is it the concept of being weakly free or strongly free that is at issue, and if the latter, why not the former instead? The answer should be fairly evident: Neo's pleasures in Matrix-1 (or some of posttransformation Beth's immediate or near-immediate

[7] On bypassing and autonomy of actional springs, see Mele (1995, pp. 166–168, 171–172, and 183–184).

intrinsic attitudinal pleasures) might all be weakly free. The new-wave surgeons may well implant unsheddable attitudes, values, etc. in their subjects but these need not, as Beth's case highlights, compromise responsibility-relevant *control*. It is not the lack of weak freedom that explains why Neo's life is not so highly good in itself for him. Rather, his life fails to be highly intrinsically valuable for him because his pleasures are not strongly free.

4.6 Freedom, Desert, and Value

Revert to the question of why free pleasures are more intrinsically valuable than otherwise similar unfree pleasures. I ventured that posttransformation Beth is not morally praiseworthy or blameworthy for (at least) her first few decisions and overt actions (though some disagree with this assessment). She is not morally responsible for these things because they are derivatively inauthentic; these decisions and actions express causal springs that are inauthentic. Just as a person may deserve overt praise or blame, so a person may deserve pleasures or displeasures or may not deserve these things. A person may deserve pleasures (or displeasures) for any number of reasons. Of special interest for our concerns are instances in which a person deserves to receive a pleasure in virtue of performing an action such as giving alms to the poor.

Reflect on the following case which suggests that the intrinsic value of freedom—the value of free intrinsic attitudinal pleasures—is merely a consequence of a more general thesis about the intrinsic value of deserved pleasures. This is roughly the thesis that deserved intrinsic pleasures (or intrinsic displeasures) of a certain sort are better than otherwise similar pleasures (or displeasures) that are not deserved.

Imagine that Beth* is as "close" to Beth as possible apart from not being manipulated. Confine attention to a segment of Beth*'s life, *sl**. First assume that for each act that Beth* performs at some time in this segment, victimized Beth, at this time, performs an act type- or near type-identical to this act in the "corresponding segment," *sl*, of her life; and for each act that Beth performs, at some time, in *sl*, there is a type- or near type-identical act that Beth* performs at that time in *sl**. So, at a time *t*, Beth* performs an act in *sl** if and only if Beth, at *t*, performs a type- or near type-identical act in *sl*. Assume, second, that in life segment *sl**, Beth* receives intrinsic attitudinal pleasures and intrinsic displeasures, each of which she deserves to receive. So in this temporal segment, it is false that on any occasion Beth* deserves to receive displeasure but receives pleasure then instead; and it is false that on any occasion, Beth* deserves to receive pleasure but receives displeasure then instead. In this segment of her life, there is a perfect "fit" between the pleasures (and displeasures) she receives and the pleasures (and displeasures) she deserves to receive. Assume, third, that all the pleasures that Beth* receives in this segment are pleasures that Beth* deserves to receive in virtue *of performing actions*, such as giving alms to the poor. (Corresponding things are true concerning the displeasures that she receives.) Assume, fourth, that for each (episode of) pleasure that Beth* receives in life segment *sl**, there is a "corresponding pleasure" that Beth receives

in life segment *sl*, a pleasure that she receives upon performing an act that is type- or near type-identical to the one Beth* performs, this act of Beth*'s being the reason for Beth*'s receiving the apt episode of pleasure (or displeasure) that she receives.

If it is plausible that Beth is not morally praiseworthy or blameworthy for any action that she performs in life segment *sl because* all these actions are derivatively nonautonomous, it is *just as plausible* that she does not deserve to receive any of the pleasures (or displeasures) that she receives upon performing the apt, derivatively nonautonomous act. A principle that captures this connection between desert of pleasure (and displeasure) and autonomy is this:

> PDA: If an agent receives a particular pleasure (or displeasure) upon performing a derivatively nonautonomous act, A, which is an instance of pleasure of the same sort of pleasure that the agent would have deserved to receive in virtue of performing an act as "close" as possible to A with the exception of being derivatively autonomous, then the particular pleasure (or displeasure) that this agent receives is not deserved.

Confining attention to life segments *sl** and *sl*, we said that for each pleasure that Beth* receives in *sl** because of performing a derivatively autonomous act, there is a corresponding pleasure that Beth receives in *sl* upon performing a derivatively nonautonomous type- or near-type identical act (and similarly for displeasures). We may say that each of the pleasures that Beth* receives is a positive desert-based pleasure: she deserves the pleasure—she has positive desert base—and she receives it. Analogously, each displeasure that Beth* receives is a negative desert-based displeasure: she deserves the displeasure—she has negative desert base—and she receives it. In contrast, each of Beth's pleasures and displeasures (in life segment *sl*) is neutral desert-based: Beth does not deserve any of them but still receives them. Axiological principles AXP1 and AXP3 entail (together with the germane facts of the case) that each of Beth*'s positive desert-based pleasures is better than each of Beth's corresponding neutral desert-based pleasures. Principles AXP2 and AXP3 entail (together with relevant facts) that each of Beth*'s negative desert-based displeasures is less bad than each of Beth's corresponding neutral desert-based displeasures.

Let us remind ourselves of the part of *assumption value* that is of concern to us: subject's desert can have an impact on well-being by way of having an impact on the value of episodes of intrinsic attitudinal pleasures and displeasures. We can now conclude that Beth*'s life is better in itself for Beth* than is Beth's life for Beth.

Collecting some salient results, deserved pleasures are (other things being equal) intrinsically better than pleasures that are not deserved. The global induction case involving Beth strongly suggests that, regarding some pleasures and displeasures, if they are deserved, then they derive from authentic springs. Which pleasures (and displeasures) are pleasures of this sort? So far, I have given a very rough characterization of these pleasures as pleasures that one deserves in virtue of performing actions, and I have called them "action-based pleasures." More precision is called for. To help isolate this class of pleasures, we remind ourselves that there are several different desert bases—there are different factors that influence the extent to which a given person deserves a certain pleasure or displeasure. Reasonable candidates include deficient past receipt, innocent suffering, moral worthiness, conscientious

effort, performance of germane action, and legitimate claims.[8] Suppose, in virtue of being a person, you deserve a certain amount of pleasure, but as time goes by, through no fault of your own you get very little of this pleasure. Because of this deficient past receipt, you deserve appropriate compensation. Or imagine that an innocent person who does not deserve to suffer has been mugged. In virtue of her innocent suffering, she deserves compensatory benefits. Next, consider a case in which there are two potential recipients of a good, alike in all other relevant respects save that one of them has been morally good (she has performed many morally good deeds, for instance), whereas the other has been bad. Barring other complicating factors, the person who has been morally worthy deserves the good or more of it. Suppose, however, that the sole reason why this person has been good is that she has been manipulated into performing good deeds by the new-wave neurosurgeons who worked on Beth. In this variation of the case, we would not think that the person who has so acted *deserves* the good. For her good deeds stem from actional springs that are inauthentic. A similar moral comes to light when we consider conscientious effort. If one student has worked long and hard to succeed in a course but another has wasted the semester, other things being equal, the hard worker deserves more to enjoy the rewards of success. If, however, we were to discover that the hard worker's efforts were due solely to apt manipulation, we would not think that he was so deserving; again, his efforts issued from antecedents of action that were inauthentic.

We may draw a distinction between *action-implicating* and *nonaction-implicating* desert bases. Roughly, an action-implicating desert base, such as conscientious effort or moral worthiness (where the latter is to be understood fundamentally in terms of performing *actions* in virtue of which one is morally good), implies that the factors that influence the extent to which a person deserves some good are germane actions (mental or otherwise) of the potential recipient. A nonaction-implicating desert base, such as being a person, is a desert base that is not an action-implicating one. We may say that an agent deserves a pleasure (or displeasure) on the basis of some desert base if and only if in virtue of that desert base, it is distributionally appropriate for this agent to receive the pleasure (or displeasure). Now we may refine the notion of *something's being an action-based pleasure* in this way: an action-based pleasure (or displeasure) is a pleasure (or displeasure) that, if deserved, is deserved on the basis of an action-implicating desert base. We may finally enunciate the following principle:

> *Deserved Action-Implicating Pleasures (DAIP)*: If an agent deserves a pleasure (or displeasure) on the basis of an action-implicating desert base, then that pleasure (or displeasure) causally arises—"it has its basis"—in springs of action that are authentic.

In sum, each action-based pleasure and displeasure that is deserved is free (in that each such pleasure and displeasure is derivatively authentic). Each deserved pleasure (and displeasure) is better than an otherwise similar pleasure (and displeasure) that is not deserved. (The pleasure is made better because deserved; the displeasure

---

[8] This list is, of course, not meant to be exhaustive and some may take issue with whether some of the items on this list should be on it.

is made less bad because deserved.) If these two things are true, freedom affects the values of pleasures and displeasures. If freedom affects the values of pleasures and displeasures, and subject's desert can have an impact on well-being, then freedom can have an impact on well-being. We have assumed that subject's desert can have an impact on well-being. It follows that freedom, too, can have an impact on well-being.

4.7 Freedom and the Value of Action-Based Pleasures

With the Beth/Beth* case in mind, perhaps one might now attempt to account for the enhanced value of free pleasures in this way: assume that deserved pleasures are more valuable for well-being than pleasures that are not deserved. If this is so, and if a pleasure cannot be deserved unless it is free, then free pleasures enhance well-being. Thus, the value of free pleasures *derives* from the value of deserved pleasures. Freedom does not always enhance the value of pleasures. But when it does, it does so via considerations of desert. An hypothesis that captures the core of this view is the following.

H1: When a free pleasure is better than an otherwise similar unfree pleasure, it is better only if this pleasure is actually deserved.

H1 is seductive but it is not correct. We may spin the tale of Neo in Matrix-1 in a way in which it is clear that none of Neo's intrinsic attitudinal pleasures is deserved and none is not deserved; each is gratuitous (for simplicity, imagine that Neo takes no intrinsic displeasure in any state of affairs). Compare Neo's life with the life of his twin. This life is otherwise similar to Neo's save for its being "real." Notably, we assume that the objects of the twin's pleasures are all just as he takes them to be, the pleasures are gratuitous, and these pleasures are free in that they causally derive from authentic springs of action. If H1 is true, each of the twin's free gratuitous pleasures has the same intrinsic value as each of Neo's corresponding unfree gratuitous pleasures. It would then follow that Neo's life in his slimy womb would be no better for Neo than would the life of his twin be for the twin. Again, this just rings false to me.

There is another problem with H1. A proponent of H1 is led to H1 partly on the basis of the principle that an intrinsic attitudinal pleasure cannot be deserved unless it is free. I'm inclined to reject this principle. In virtue of deficient past receipt or, perhaps, merely in virtue of being a person, you may deserve certain pleasures. In a case of this sort, it may be that you deserve the pleasures but *not* because of any action of yours. A pleasure, then, may well be deserved even though it is not indirectly free.

Consider two other competing hypotheses:

H2: A free pleasure is better than an otherwise similar unfree pleasure because freedom is intrinsically good.
H3: A free pleasure is better than an otherwise similar unfree pleasure for just

the reason or for a reason similar to the reason that (i) a pleasure in a true state of affairs is better than an otherwise similar pleasure in a false state of affairs, or (ii) a pleasure in an object that deserves to have pleasure taken in it is better than an otherwise similar pleasure in an object that does not deserve to have pleasure taken in it.

Unlike H1, both H2 and H3 imply that freedom enhances the intrinsic value of a pleasure per se. On these hypotheses, the twin's life is better in itself for the twin than is Neo's life for Neo. H2 is unacceptable to an axiological monist, such as a freedom-sensitive attitudinal hedonist, who insists that there is only one property, which is such that intrinsically good basic intrinsic value states are pure attributions of that property. If H2 is true, and it may well be, I believe that its truth does not detract from the fact that freedom can have a definite impact on well-being. This is because I do not see how any *plausible* account of welfare can deny that intrinsic pleasures and displeasures contribute, in a fundamental way, to well-being.

I am *somewhat* partial to the view that truth and "object-worthiness" can enhance the value of an episode of intrinsic attitudinal pleasure. For this reason, I am drawn to H3. Still, H3 is a bit mysterious (an understatement! one might exclaim): it fails to specify *why*, for instance, object-worthiness is value-enhancing. One cannot, of course, rule out of hand an explanation along the lines that an object's being intrinsically good accounts for its being deserving of having pleasure taken in it. In this event, would *object-worthy hedonism* fail to qualify as a welfarist form of *hedonism* because there are things besides attitudinal pleasures that are intrinsically good? I think not; I think the supposition that freedom is intrinsically good wouldn't threaten *object-worthy hedonism's* claim to being an hedonistic life-ranking axiology. A theorist may consistently hold that, say, a beautiful object is pleasure-worthy because this object is intrinsically good, its being intrinsically good does not contribute independently to *prudential* value, and its being pleasure-worthy enhances prudential value but only because its being pleasure-worthy augments the values of pertinent basics involving pleasures in pleasure-worthy objects.

I said in the opening paragraph of this work that there are many different varieties of moral appraisal. There are, for instance, morally deontic judgments of right, wrong, and obligation; hypological judgments of moral praiseworthiness and moral blameworthiness; aretaic judgments having to do with virtue and vice; and axiological judgments concerning individual well-being and the value of worlds. There is no presumption that the conditions requiring satisfaction for one species of moral judgment to be true (or "correct") are the very conditions requiring satisfaction for a different species of moral judgment to be true (or "correct"). But with respect to at least some of these varieties of judgment, it seems that there is some overlap in the conditions of satisfaction. It is generally presumed that responsibility requires freedom: no one can be morally praiseworthy or morally blameworthy for an action unless that action is free. Similarly, there is a requirement of freedom for the truth of morally deontic judgments. In a world bereft of the pertinent freedom, neither judgments of responsibility nor judgments of deontic morality are true. We may now

add that there may well be a pivotal association between judgments of well-being and freedom, on the assumption that a freedom-sensitive attitudinal hedonism that adjusts the values of episodes of pleasure and displeasure to reflect the extent to which these pleasures are deserved or not is in the right ball park *and* pertinent to well-being: free pleasures (and displeasures) derive from authentic springs of action and, other things being equal, such pleasures (and displeasures) are better than otherwise similar unfree pleasures (and displeasures).

Chapter 5
Authentic Springs of Action

A pleasure (or displeasure), if free, is indirectly so; it derives its freedom from the freedom of decisions. A decision, in turn, is free only if it causally issues nondeviantly from springs of action that are authentic. Thus, a pleasure (or displeasure) is free only if it derives from authentic antecedents. Under what conditions are our springs of action authentic and under what conditions are they inauthentic? Since I (and my coauthor) have undertaken significant parts of this project elsewhere (Haji and Cuypers 2008), in this chapter, I adumbrate aspects of this prior work on authentic springs that are of immediate significance to the axiological issues under discussion.

5.1 Authenticity and Welfare

To develop an account of authenticity, distinguish between two stages in an individual's life: the stage, roughly, of childhood prior to which one is a suitable candidate for responsibility ascriptions and the stage when or after one satisfies the agency requirements of responsibility. Some background will be helpful.

In addition to freedom and epistemic constraints, moral responsibility has agency requirements. Very young children are not responsible for their conduct partly because they fail to fulfill responsibility's agency presuppositions. To be morally responsible, one must be an agent of a certain sort, what I call a "morally normative agent." One agency requirement for responsibility is that the candidate be capable of intentional deliberative action. Such action, in turn, requires some psychological basis for evaluative reasoning. An agent's deliberations that issue in a practical judgment about what to do, which in turn gives rise to a decision or intention, involve an appraisal of reasons for or against action by appeal to what I dub the agent's "evaluative scheme." Such a scheme is made up of these constituents: (a) normative standards that the agent believes (though not necessarily consciously) ought to be invoked in assessing reasons for action or beliefs about how the agent should go about making choices. To be a fitting candidate for *moral* responsibility, the normative standards must include a set of moral principles or norms; the agent must be minimally morally competent. She must understand the concepts of rightness, obligatoriness, or wrongness, and she must be able to appraise, morally, various choices

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

or actions in light of the moral norms that are elements of her evaluative scheme. (There is no requirement that appraisals be fully considered, free of error, or even conscious.) (b) The agent's long-term ends or goals he deems worthwhile or valuable. (c) Deliberative principles the agent utilizes to arrive at practical judgments about what to do or how to act. (d) Lastly, motivation both to act on the normative standards specified in (a) and to pursue one's goals of the sort described in (b) at least partly on the basis of engaging the deliberative principles outlined in (c).

It *suffices* for an individual's being a normative agent at a certain time that the individual have at that time (i) an evaluative scheme with the requisite moral elements—the agent is minimally morally competent; (ii) deliberative skills and capacities; for example, the agent has the capacity to apply the normative standards that are elements of his evaluative scheme to evaluating reasons; and (iii) executive capacities—the agent is able to act on at least *some* of his intentions, decisions, or choices. Read condition (ii) to entail that the agent is able to engage in genuine deliberation; his deliberative activities must meet the threshold of rationality below which such activities fail to count as bona fide deliberation.

Evaluative schemes contain both doxastic and motivational components. Again, for brevity, attention is limited primarily to the latter that include desires. As a child matures, the child acquires an evaluative scheme; the child becomes a normative agent. The child's *initial evaluative scheme* is the evaluative scheme that the child initially acquires. One guiding idea is that if the motivational constituents of such a scheme are properly acquired, in a sense of "proper acquisition" to be explained, then these constituents are authentic. Over time, the initial scheme evolves; its constituents change. If we can give an outline of when the desires that are parts of an individual's initial scheme are authentic, then we can theorize that the desires of an evolved scheme are authentic provided that they causally result from modifications to the individual's initial scheme that are acceptable, again in a sense of "acceptable" to be explained. Here, I provide only an outline of authenticity.

Regarding the motivational components of an *initial scheme*, such as desires or, more generally, pro-attitudes, I propose that these motivational elements, if authentic, are not authentic in their own right. Rather, they are relationally authentic: they are authentic only relative to whether later behavior that issues from them is behavior for which the morally normative agent into whom the child matures is morally responsible. More specifically, with respect to pro-attitudinal components or would-be components of the initial evaluative scheme that the child acquires, I suggest the following. *A pro-attitude or its mode of acquisition is inauthentic if that pro-attitude or the way in which it is acquired subverts moral responsibility for behavior, which owes its proximal causal genesis to the pro-attitude (typically in conjunction with other springs of action), of the morally normative agent into whom the child develops.* Subversion of moral responsibility would occur as a result of either epistemic, control, or other necessary requirements (*independent*, of course, of agency presuppositions) of moral responsibility being thwarted. In this sense, at the prenormative scheme stage there is nothing like "authenticity per se" of pro-attitudes but only relational authenticity: "authenticity-with-an-eye-toward-moral responsibility" or "responsibility-relative authenticity." I revert to this point below.

Toward outlining an account of initial scheme authenticity, ponder these examples. I said that to be morally responsible for an action, an agent must have elementary moral concepts, such as those of wrong or obligation, and she must be able to appraise morally (even if imperfectly) decisions, actions, consequences of action, etc. in light of the moral norms that are partly constitutive of her evaluative scheme. Imagine that a child, Simon, is trained in such a way that he lacks knowledge of the relevant moral concepts with the result that he is not even *minimally* morally competent. Then lack of instilment of the appropriate moral concepts is responsibility-subversive. Or consider instilment in Simon of a pro-attitude or disposition, the influence of which on his behavior he cannot thwart. Instilling such a pro-attitude would presumably undermine responsibility for later conduct arising from that pro-attitude by undermining the control moral responsibility requires. Or suppose instilled in Simon is a powerful disposition always to act impulsively. Here, again, we would not want to hold Simon morally responsible for much of his later impulsive behavior. Or, finally, consider an interference that prevents Simon from engaging in critical self-reflection. This may subvert Simon from being morally responsible for some of his later behavior, on occasions of choice, by significantly narrowing the range of Simon's options, alternatives he may have considered had he acquired "normal" habits of critical self-reflection.

In each of these cases, Simon acquires a certain pro-attitude. The pro-attitude (or the way in which it is acquired) is such that behavior that it later causes (when Simon becomes a normative agent) is not behavior for which Simon will be morally responsible. His lack of responsibility for this behavior will be due to the fact that the pro-attitude, or the manner of its acquisition, ensured that some necessary condition of responsibility (such as the control condition) is not satisfied. The pro-attitude in question, for example, might be a desire that is irresistible. The action (if any) that results from such a desire at a time when Simon has become a normative agent will not be one for which Simon is praiseworthy or blameworthy. This is simply because Simon won't exercise the control that responsibility requires in performing an action that issues from this desire. The irresistible desire is not inauthentic in its own right. It is inauthentic (at the child's or prenormative agent's stage) *only because* it will thwart later moral responsibility. In this way, it is relationally inauthentic.

In sum, let "interference" be a general term for things like suppression of innate propensities, or implantation of certain dispositions, or deliberate lack of instilment of various pro-attitudes. Some interferences are incompatible with the agent being morally responsible for his subsequent behavior, which issues from these interferences. Such interferences subvert later moral responsibility (as I will say), while others do not. I propose that the subversive ones are *responsibility-wise inauthentic*. Specifically, imagine an agent like a young child who does not yet have an initial scheme. Such an agent's having a pro-attitude is responsibility-wise inauthentic if her having this pro-attitude as a result of instilment subverts her being morally responsible for behavior to which it gives rise; owing to having this pro-attitude, she is not responsible for conduct that stems from it. "Stems from" requires analysis that shall not be undertaken here save for the following. A causal theory of action (which I endorse) assumes that actions causally arise from desires, or desire/belief pairs, or

a cluster of psychological elements. On this theory, when an action issues from a certain desire (as opposed to another), this desire (as opposed to the other, typically together with other actional elements) is causally implicated in the production of the action. I presuppose whatever account of "issues from" that causal theories of action presuppose. Setting aside agency requirements of responsibility, if interferences of the sort at issue subvert later moral responsibility, they will do so by subverting *other* requirements of responsibility, such as epistemic or freedom requirements.

We have, so far, limited discussion to responsibility-relevant authenticity of the objects of instilment such as dispositions, or pro-attitudes in general. What about modes of instilling such things; are some responsibility-wise authentic and others not? Assume that to ensure prevention of subverting moral responsibility for later behavior, it is necessary to instill in the child the disposition to be moral. Different modes of instilling this disposition could affect responsibility-relevant authenticity of this very disposition itself. For example, suppose that given the mode of instilling the moral disposition in Simon, he finds that he cannot refrain from doing what he perceives to be morally right. Perhaps the disposition was "beaten into" Simon, or instilled via shock "therapy." On occasions of choice, he is stricken with inward terror at the faintest thought of not doing what he deems moral. Simon would not be morally responsible for much of his later behavior because the mode of instilling the moral disposition subverts responsibility-grounding control. Modes of instilling pro-attitudes (habits, dispositions, etc.) are responsibility-wise not truly one's own if the modes subvert responsibility for later behavior. Again, leaving aside agency demands of responsibility, if these modes of acquiring pro-attitudes undermine later moral responsibility, they will do so by subverting one or more of responsibility's requirements.

We may now formulate a general principle about the authenticity of initial evaluative schemes. I have suggested that the having of some pro-attitudes (dispositions, etc.) may be required to ensure moral responsibility for later behavior—having them ensures that necessary conditions other than agency conditions of moral responsibility can (later) be satisfied by the agent or by her behavior that stems from them. I have also suggested that the having of some pro-attitudes (dispositions, etc.) is incompatible with moral responsibility for later behavior, which issues from them; having these pro-attitudes precludes satisfaction of necessary conditions for moral responsibility other than agency conditions, such as epistemic or control conditions. Such pro-attitudes are inauthentic. Lastly, I have suggested that some modes of instilling pro-attitudes (dispositions, etc.) are incompatible with moral responsibility for later behavior; such modes of instilment subvert later responsibility by thwarting satisfaction of necessary conditions of responsibility other than agency conditions. Such pro-attitudes are also inauthentic. Limiting attention to the motivational or pro-attitudinal elements of an initial scheme, I propose this principle as one that governs responsibility-relative authenticity of *initial* schemes of "developing agents":

Authenticity-1: An agent's initial evaluative scheme is responsibility-wise authentic if its pro-attitudinal elements (i) include all those, if any, that are required to ensure that the agent will be morally responsible for her future behavior; (ii) do not include

any that will subvert the agent's being responsible for future behavior that issues from these elements; and (iii) have been acquired by means that, again, will not subvert the agent's being responsible for her future behavior.

I turn to amplification of this account of initial scheme authenticity by responding to the objection that *Authenticity-1* is either *empty* or *circular*. Regarding the former, some may think that *Authenticity-1* is empty because I have left open what *all* the other requirements or dimensions of responsibility are, and of the dimensions that I have acknowledged—the epistemic and control dimensions—I have not said in what these consist. So the implication of *Authenticity-1* that if (during the pre-initial scheme stage) a pro-attitude, such as a desire, or its mode of acquisition, undermines future responsibility for behavior that issues from it, that pro-attitude is inauthentic, is empty.

In response, setting aside responsibility's agency presuppositions, there is widespread agreement that responsibility has additional requirements. There is (general) concurrence, for example, that there are epistemic and freedom demands on responsibility. That responsibility has various requirements cannot be denied even in the face of disagreement regarding precisely *what* these requirements might be. Concerning a putative particular requirement, such as the epistemic one, theorists may well disagree on the *substantive account* of this requirement. Yet, again, though, free will theorists do *not* deny that there is *some correct rendition* of this requirement (whatever it may turn out to be). Reconsider the claim that a pro-attitude acquired during the pre-initial scheme stage is inauthentic if behavior that issues from it is behavior for which the agent is not morally responsible by virtue of this pro-attitude's undercutting one or more of responsibility's requirements. On the assumption that there is a *correct* account of what the requirements of responsibility are (with the exception, again, of agency requirements), and that there is *correct* account of what each of these requirements consist in, this claim is decidedly *not* empty.

Regarding the concern of circularity, the consideration that there is a fact of the matter about what the requirements of responsibility are, and that there is a fact of the matter about what each of responsibility's requirements amounts to, also vindicates the view that *Authenticity-1* is not circular. I do not explicate the notion of initial scheme authenticity by covert appeal to the authenticity of the scheme's constituents. I reemphasize that on my view, during the pre-initial scheme stage, there is nothing like authenticity per se of pro-attitudes acquired during this stage. The authenticity of pro-attitudes is forward-looking, specified by way of a relation between the having of a pro-attitude at a time when the child is not yet a normative agent, and behavior that issues from this pro-attitude at a time when the child has turned into a normative agent. Second, I account, specifically, for the authenticity in question in terms of whether an agent's behavior that owes its proximal causal genesis to this pro-attitude (typically, in consort with other actional elements) is behavior for which the agent is morally responsible. I ask whether the pro-attitude subverts responsibility for such behavior by subverting one or more of the requirements (other than agency ones and ipso facto other than authenticity ones) of responsibility. Assuming

that there are such requirements (whatever they may turn out to be), and that there is a *correct* account of what these requirements consist in, I see *no* circularity in *Authenticity-1*.[1]

Assume that Simon has acquired an authentic initial evaluative scheme. Evaluative schemes are not static but dynamic; typically they evolve. So, for instance, Simon can renounce values formerly cherished and acquire new ones; he might come to question his belief that moral decisions should conform to the teachings of his religion and adopt a utilitarian outlook; or he might give up some of his deliberative principles such as the principle that he review his decisions frequently before implementing them. Some modifications or changes in one's evaluative scheme may be perfectly compatible with preserving responsibility-relative authenticity, whereas others will subvert authenticity. To distinguish between the two sorts of changes, we require a conception of *acceptable* modifications.

As one's evaluative scheme is comprised of doxastic and motivational constituents, changes in one's scheme can involve changes in one or both. The general rule for acceptable modifications in either of these types of constituents is straightforward: the modifications must be made under one's own deliberative control. With respect to changes in pro-attitudes such as desires, instilled ones or newly acquired ones are acceptable as long as the actions, if any, to which they give rise are ones over which the agent has responsibility-grounding control and the changes are initiated by the agent's exercising (or engaging) her initial scheme. The changes occur as a result of capacities, such as deliberative ones, that the agent (substantially) has in virtue of elements constitutive of her authentic scheme.[2]

To elaborate, Mele plausibly proposes that most normal, healthy human agents have the following capacities in some measure: the capacity to modify the strengths of their desires in the service of their normative judgments, of aligning their emotions with relevant judgments, of mastering motivation to produce or sustain beliefs that would violate their principles for belief-acquisition and belief-retention, of rationally assessing their values and principles, of identifying with their values and principles on the basis of informed critical reflection, and of modifying their values and principles should they judge that this is called for (Mele 1995, pp. 166–172; 183–184). I do not see how it is possible to possess these capacities without having an evaluative scheme. One cannot assess values and principles, for instance, without embracing a set of normative principles. These principles are causally and, ideally, nondeviantly engaged, for instance, in one's appraisal of a thought that strikes one upon witnessing a disastrous event. If one cares deeply about another, one must be able to adjust or adapt one's emotions: appropriately favorable emotions in response to one's belief that the cared-for has fared or will fare well, and a range

[1] I (and my co-author) entertain and respond to other objections in Haji and Cuypers (2008).

[2] I allow for cases in which, through a series of past steps over which one has responsibility-grounding control, one deliberately instills in one a pro-attitude (or a cluster of such attitudes), which will give rise to actions over which one will lack such control. A person desperate to quit smoking, for example, may have implanted in her an irresistible desire to avoid cigarettes.

of unfavorable emotions in response to one's belief that the cared-for has fared or will fare poorly. One cannot do so without engaging elements of one's evaluate scheme. Here, again, there will presumably be an appropriate causal story to be told about how one brings one's relevant emotions (the having of which themselves depends upon constituents of one's evaluative scheme) in line with one's pertinent judgments. Consonant with what Mele says (see, e.g., Mele 1995, pp. 166–172, 183–184; and 2006, pp. 166–167, 170), I propose that an agent's evaluative scheme is not engaged in, for instance, acquiring a pro-attitude, if the acquisition of this pro-attitude bypasses all of the agent's capacities of deliberative control. The modification to one's evaluative scheme resulting from its supplementation with this pro-attitude is not an acceptable one. If, in acquiring a pro-attitude, one manifests deliberative control and, thus, in this sense "engages one's evaluative scheme," the degree of deliberative control that one exercises will be a function of a number of factors. These include factors such as whether the deliberative process involves certain sorts of inefficiency and irrationality—for example, various sorts of selective biasing, and the coming to mind, while deliberating, of irrelevant considerations or akratic influences.

Assuming an appropriate account of authenticity for doxastic elements, I now formulate the following sufficient condition of scheme responsibility-relative authenticity of "developmental agents" like us.

Authenticity-2: If an agent's evaluative scheme at a time is either her initial responsibility-wise authentic scheme at that time, or is an evolved responsibility-wise authentic scheme of hers at that time—it is a scheme resulting from *acceptable* modifications to a scheme that she possesses prior to that time that is responsibility-wise authentic—then her evaluative scheme is responsibility-wise authentic.

Many more details regarding the authenticity of both initial and evolved schemes need to be supplied in a comprehensive account of authenticity. The sketch I have provided, though, should be adequate for the task at hand.

Reverting to global induction cases, these cases as standardly presented, such as the Ann/Beth case, raise concerns not about normative agents in the making but about fully formed normative agents with evolved evaluative schemes. Beth, in Mele's example, is such an agent. Global manipulation has the effect of subverting normative agency in this way: pro-attitudinal and doxastic components of the individual's evaluative scheme are "replaced" by a different set. The replacement is not accomplished under the individual's own steam but occurs as a result of some process that totally bypasses the agent's capacities of deliberative or reflective control (cf. Mele 1995, pp. 166–168, 171–172, 183–184). Victimized Beth's desires are, thus, not authentic.

5.2 A Comparison with Noggle's Account

It will be helpful to compare this relational account of authenticity with what Robert Noggle has to say about authentic springs at the prenormative agent stage of devel-

opment. Introducing some terminology, Noggle directs our attention to the term "self" according to which the person's self does not include those "internal but phenomenologically alien forces that may afflict her and threaten her personal autonomy" (Noggle 1999, p. 88). Noggle adds:

> The adjective "authentic" is commonly applied to elements of the person's psychology that are part of or produced by this true or real self. Thus, to say that an impulse is not authentic is to say that it does not lie within that part of a person's psychology that must be in charge if she is to be genuinely autonomous (or that must be the source of her actions if they are to count as autonomous) (p. 88).

Noggle proposes that the "authenticating self first arises gradually" (p. 99). He affirms that if a self is to "ground assertions of authenticity," it must have a stable, orderly belief system and preference structure, and it must have the psychological wherewithal necessary to allow it to reflect upon and revise those beliefs and desires (p. 99). He claims that a person's desires and beliefs are structured around a core, consisting of "those beliefs that constitute her most basic cognitive organizing principles and fundamental assumptions and convictions, together with the desires that constitute her deepest, most significant goals, concerns, commitments, and values" (pp. 99–100). These core attitudes, comprising a relatively stable framework for the agent's psychology, play a fundamental role in making the person who she is and "giving shape to the rest of her psychological elements" (p. 100). Once a person has acquired such a core self, she has the ability to adjust and revise the doxastic and motivational elements of its core. Noggle elucidates:

> Often, such changes are, to a large degree, "internally motivated" in such a way that they seem to be *intelligible reflections* of the contents of the core attitudes. Such changes resolve contradictions, inconsistencies, or other kinds of tension among core attitudes or between a core attitude and persistent information about oneself or the outside world. When changes to the core attitudes are of this kind, the self evolves according to its own internal logic—its own contents determine whether and how it is to change in response to new information, internal conflicts, and changing conditions. . . . When psychological changes happen this way, it seems correct to say that the new configuration of the self is an authentic continuation of the previous configuration. On the other hand, a psychological change—especially a change to the core attitudes—that does not occur in this way produces a new configuration that is not an authentic continuation of the previous one. (pp. 100–101)

Turning now to the issue of pressing interest to us, how does such a first self arise? Noggle writes:

> Infants and very young children do not yet have the two key psychological ingredients for the kind of self that we are supposing is the determiner of authenticity. The infant's cognitive structures and capacities are unformed, and her motivational system consists mainly of unstructured biological drives. As the child grows, she begins to develop cognitive structures around which she will organize her beliefs, as well as the stable concerns, attachments, and goals that will provide structure to her motivational system. Together, these will gradually coalesce to form the core of her self. The earliest core desires, as well as the initial elements of the child's cognitive conceptual scheme, arise via processes that would be considered authenticity undermining if they were used to implant beliefs and desires into an adult. Such processes apparently include operant, aversive, and classical conditioning; role model imitation; blind obedience to and subsequent internalization of behavioral norms; uncritical acceptance of propositions on the authority of parents and teachers; and so on. Out of a

> seemingly unpromising beginning—a sort of chaotic psychological "soup"—the child's self gradually emerges as her cognitive and motivational systems develop the kind of structure and stability and the rational and reflective capacities necessary for the existence of a coherent and stable self that can be the source of authenticity. (p. 101, note omitted) [3]

In short, Noggle's answer to whether a desire or a belief that is "beaten" into a child at the prenormative agent stage is authentic is "yes":

> If we accept a self-referential condition of authenticity, an element is authentic to a person just in case it bears the right relation to her true self. Before the self initially arises, there is no other self for the initial self to bear any authenticity-grounding relation to. . . . When that initial self forms, it is the only self that there is. Sadly, that initial self is the only game in town, so to speak. Now if we ask whether some *element* of that initial self is authentic, then the answer simply has to be "yes." After all, the element belongs, *ex hypothesi*, to the only self that exists. If the self is fully formed and the elements are related to it in the right way (with the right way depending on what theory of authenticity we finally adopt), then that is all there is to their being authentic (p. 103).

While there is much with which I agree in Noggle's account, I believe that the implication of his account that any doxastic or pro-attitudinal element in the child's psychological repertoire (at the stage during which the child is not yet a fully developed normative agent) qualifies as authentic is seriously flawed. I have already supplied examples that motivate rejection of this implication. If a religious, fundamentalist leader successfully manages to instill in the young who are forced to attend his sermons desires that are irresistible, then on the relational view (but not on Noggle's) these desires are (relationally) inauthentic. The inauthenticity of such desires (and other actional elements) has implications for the authenticity of various other desires (and actional elements) that the adolescent into whom the child develops will come to possess. Consider the following. Suppose relationally inauthentic desires (including higher-order ones) and beliefs, instilled at the pre-initial scheme stage, are later implicated in the "fully formed" self's deliberations regarding changes to her core attitudes. Then, it seems, these deliberations inherit the taint of inauthenticity owing to some or all of the primary "inputs"—the instilled elements—to deliberation being inauthentic themselves. It should come as no surprise, then, that the "output" of these deliberations—"altered" beliefs or desires—may themselves not be authentic. This could be so even though the deliberations of the present self involve no empirical errors or logical confusions.

5.3 Some Objections and Responses

I now respond to three objections to the relational account of authenticity. The first questions the assumption that pre- and postmanipulated Beth are identical. It may be rejoined that the Ann/Beth case could be interpreted as follows: insofar as a person's identity is *constituted* by her values (or, more generally, her evaluative scheme), a wholesale change in values changes the person's identity. The posttransformation

[3] Noggle expands on this in Noggle (2002).

person, as a result, would be "Beth-2" (or, more accurately, "Ann-2," given that she would be exactly similar to Ann). But then this "newly born" person would be responsible for her actions, having just received an authentic initial evaluative scheme, albeit in an unconventional fashion. Indeed, if it is one's evaluative scheme that matters for both authenticity and autonomy, this seems by far the more plausible move (i.e., to admit that agents are identified with their evaluative schemes, so global induction produces new identities).

I stress that it is not all of Beth's values that are replaced; nothing in the thought experiment requires such total change. Maybe premanipulated Beth is kind and generous, intellectually meticulous, and a connoisseur of food and drink. The manipulation need not affect any of the values associated with these things. Postmanipulated Beth need have no false pseudo-memories either: she may be genuinely surprised by the change in her. Moreover, she may rationalize this change. She might believe that it is time that she makes a lasting, worthwhile intellectual contribution and that she can only do so if she dedicates herself wholeheartedly to philosophical pursuit. With these stipulations, the pre- and postmanipulated agents may be strongly psychologically connected, in Parfit's sense of "connected" (Parfit 1984, p. 206). At stake is whether postmanipulated Beth is morally responsible for actions that express unsheddable engineered-in values, pro-attitudes, and beliefs, not those that the secretive manipulation leaves untouched.

What of the suggestion that if it is one's evaluative scheme that "matters for both authenticity and autonomy," then it ought to be that agents are to "be identified with their evaluative schemes?" One concern with this suggestion is that it implies that adult Ann is not identical to neophyte Ann. A second primary concern is that the suggestion conflates ontological categories: whereas an agent is a substance, an evaluative scheme is not. A substance cannot be identical to a nonsubstance. Maybe there is some sense of "identity" pertinent to responsible agency that is not the sense of "identity" in a remark such as "adult Ann is the same person as baby Ann." The objection, though, concerns *personal* identity: are pre- and postmanipulated Beth the same person? It makes little sense to claim that in this sense of "identity," premanipulated Beth is identical to her evaluative scheme, and then argue that pre- and postmanipulated Beth are not the same persons because their evaluative schemes differ.

The second objection concerns actions for which an agent is allegedly morally responsible but which express changes in outlook that seemingly occur as a result of bypassing the agent's minimal capacities of reflective control. The actions in question may, for example, be expressive of certain "new values," the acquisition of which is not under the agent's control. We are invited to think of a modification of Frankfurt's volitional necessity cases, in which, say, many of one's values are changed by an overwhelming care or commitment one has to some object. The phenomenon is defined as a case in which one cannot do otherwise, in a very important sense, and, further, as one in which one would not *want* to do otherwise, given that this would involve a betrayal of something one cares deeply about. So we can imagine one's values changing in response to some overwhelming care (as in loving someone), and even though one cannot prevent the change, one would not *want* to

prevent it. It is far from settled whether such cases involve the agent's "exercising" her initial scheme in any active sense. In addition, there is this sort of case: suppose a selfish person's evaluative scheme undergoes sudden and drastic change through her witnessing some catastrophe—the devastation caused by the earthquake in China, say—and that the nature of her actions is accordingly drastically changed. We would be reluctant to declare her not morally responsible for these actions, even though the change in her outlook seems not to have been carried out under her own steam.

In response, taking the "necessities of love" as paradigm instances of volitional necessity, according to Frankfurt, it appears that the volitional necessity to which a lover is subject involves his being irresistibly *motivated* to act in the interest of his beloved and his irresistibly *identifying* with this motivation:

> The lover cannot help being selflessly devoted to his beloved. . . . It may seem that in this respect love does not differ significantly from a variety of other familiar conditions. There are numerous emotions and impulses by which people are at times gripped so forcefully and moved so powerfully that they are unable to subdue or resist them. . . . But irresistible forces do not invariably oppose or conflict with desires or intentions by which we would prefer to be moved. They may move us irresistibly precisely in ways that we are wholeheartedly pleased to endorse. There may be no discrepancy between what we must do and how we would in any event wish to behave. In that case, the irresistible force is not alien to us at all. (1999, pp. 135–137)

Recall the gloss on one's evaluative scheme's being engaged or exercised: we have various capacities of deliberative control. The having of these capacities supervenes upon features or constituents of one's evaluative scheme. An agent's evaluative scheme is not engaged in, for instance, acquiring a desire, if the acquisition of that desire bypasses all of the agent's capacities of deliberative control. Presumably, the motivation that the lover acquires to act in his beloved's interest is *not* motivation that fails to engage the evaluative scheme of the lover. And the lover's identifying with this motivation is, again, presumably, something that does not bypass the lover's capacities of deliberative control since, presumably, identification cannot be divorced from one's deliberative principles.[4]

As for cases involving sudden or drastic conversions, as Mele (2006, pp. 179–184) has argued, whether the agent *is* morally responsible for the pertinent conduct pivots vitally on the filling in of relevant details. On the one hand, suppose (implausibly) that the witness to the earthquake undergoes the changes that she does because, at the time of the devastation, God implants in her a powerful disposition to be charitable, the implantation mirroring the implantation of manipulated Beth's new values. In this case, it is less than obvious whether the witness *is* morally responsible for her pertinent deed. On the other hand, imagine that the traumatic event generates in the agent an insight into the human condition, which she then evaluates and which subsequently moves her to change her outlook. In this variation of the scenario, it is false that the agent's evaluative scheme is idle: her exercising capacities of deliberative control is crucial in explaining the change.

---

[4] For an extensive discussion on the concepts of activity, passivity, and volitional necessity that are relevant to this objection, see Cuypers (2000).

Finally, the third objection concerns alleged equivocation on the term "autonomy". It may be put to us that the reason we think Beth's autonomy has been compromised is that the manipulator has indeed violated her autonomy in the *moral* sense of the word—she gave no *consent* to the interference—but he did not necessarily violate her autonomy in the *responsibility-providing* sense of the word. So we can certainly agree with Mele that Beth's moral autonomy has been undermined, but if we err in inferring from our agreement on that term that her responsibility was then undermined, we may very well be making a mistake: her failure to consent to the interference does not render her nonautonomous in the sense (ostensibly) required for moral responsibility. [5]

Like Mele, I take the global induction of the sort exemplified in Beth's case to be responsibility-undermining but the objector does not. Presumably, though, this objector regards *some* forms of manipulation or treatment, or *some* sorts of interference, as responsibility-subverting. Assume that Hal, in the absence of his consent, has been subject to this sort of treatment. Assume, further, that the objector's opponents do not regard this type of treatment as threatening responsibility. It would be ineffective (to say the least) for such opponents to argue against the objector in this way: "As Hal did not consent to the treatment, his moral autonomy has been violated. But it is a mistake to infer from this that Hal is not responsible for his pertinent behavior." Our objector would presumably not be moved by this sort of argument. She may well agree with the imagined opponents that the treatment violates Hal's moral autonomy. But further considerations would be required to persuade her, contrary to what she believes, that the treatment itself is not responsibility-undermining and hence that Hal is responsible for his pertinent behavior.

Reverting to Beth's case, it is open to the objector to supplement the "no consent argument" with additional factors that tell against Beth's not being morally responsible. For instance, the objector might appeal to "sudden conversion cases" or "sudden change of outlook" cases to convince us that in these cases, despite undergoing the sudden changes, the agents are still morally responsible for their germane behavior, and then add that Beth's case is not relevantly different from these cases. My response to the second objection, though, casts doubt on whether the sudden change cases can turn the trick. When they are interpreted in a way in which they are analogous to Beth's case, it is contentious whether the germane agents are morally responsible for their pertinent acts. When interpreted differently in the manner suggested, the germane agents may well be responsible.

5.4 Authenticity and Well-Being: an Objection

The relational account of authenticity may prompt the following objection to a version of freedom-sensitive hedonism that implies that free intrinsic attitudinal pleasures are better than otherwise similar unfree pleasures and that free intrinsic

[5] For the development of this sort of objection, see Arpaly (2003, pp. 126–128).

displeasures are not as bad as otherwise similar unfree displeasures. Reverting to the Beth/Beth* case, suppose (manipulation-free) Beth* receives some pleasures—action-based ones—because, roughly, of certain actions that she performs. Imagine that each of these pleasures is free; so each derives from authentic actional springs, such as authentic desires. Imagine, further, that (manipulated) Beth also receives action-based pleasures as similar as is possible to the ones that Beth* receives save that each of these pleasures derives from inauthentic springs, springs that are otherwise as "close" as possible to the ones that give rise to each of Beth*'s pleasures. Perhaps these springs were all engineered-in against Beth's wishes. Beth has no idea whatsoever that these pleasures derive from inauthentic springs. Finally, assume that neither Beth* nor Beth take intrinsic displeasure in anything over the relevant period. The objection is that, contrary to the version of freedom-sensitive hedonism under scrutiny, each fares *just as well*. So authenticity of actional springs really has nothing fundamental to do with well-being.

Again, I am not sure how convinced one should be by this objection. Suppose one believes that truth does affect welfare. One is drawn, for example, to the view that the deceived businessman does not fare as well as the businessman who is not deceived even though their lives are internally indiscernible—they "feel" the same from the "inside." Then it seems that the sort of objection in question, advanced to a proponent of the view that truth (relevantly) influences well-being, would lose much of its luster. Indeed, against the objection, we might remind ourselves that well-being concerns the good of the *agent*. If there is a transparent sense in which the intrinsic pleasures of an agent are not "really the agent's own," we would be hard-pressed, or so it may be urged, to suppose that the *agent* is better off as a result of receiving these pleasures. There *is* something compelling about the thought that "inauthentic pleasures"—attitudinal pleasures that derive from inauthentic springs—are not "really" one's own pleasures and, hence, should contribute less to personal well-being than should otherwise similar pleasures that are authentic. We will revert to this sort of objection (in Chapter 11) when we discuss preferentism.

5.5 Authenticity and Well-Being: Another Objection

Reflect on another related objection to the view that the freedom of pleasures (and displeasures) can affect personal well-being. The objection proceeds indirectly by first inviting critical attention to whether truth and object-worthiness do indeed bear on welfare. Regarding the former, suppose you suspect that your spouse is in an adulterous relationship. Failing to confront the evidence candidly, you engage in self-deception with respect to your spouse's pertinent behavior. Assume that you are happy to carry on in this fashion because you do not, in the end, care about the truth. You continue to take intrinsic pleasure in the false state of affairs that your spouse is faithful. It may now be proposed that if it is really the case that one is indifferent to truth, there is little reason to settle for the view that, other things being equal, truth enhances the value of an episode of intrinsic attitudinal pleasure. Similarly, with respect to object-worthiness, suppose Spike* has no qualms about whether the

painting in which he takes pleasure is the real thing or a good look-alike. Even assuming that some objects deserve to have pleasure taken in them whereas others do not, Spike* doesn't really care about object-worthiness. Object-worthy hedonism implies that the value of the state of affairs, *Spike*'s taking intrinsic attitudinal pleasure at a certain time (to a specified degree) in the painting*, is not as high as what its value would have been had Spike* taken intrinsic attitudinal pleasure in the genuine item. But, again, the objector might claim that this result is erroneous. Maybe there is some value in taking pleasure in objects that deserve to have pleasure taken in them, but this value is not or does not enhance *prudential* value. Finally, the objector might further propose that there is a relevant similarity among truth, object-worthiness, and freedom: perhaps there is value to a person's intrinsic pleasures being free—in the sense of being derivatively authentic. But if one does not care about authenticity, why should it be thought that such value is pertinent to *welfare*? Why is such value germane to a *prudentially good life*?

I am sympathetic toward this line of reasoning regarding the bearing of truth and object-worthiness on well-being. I think, though, that there is something different going on with authenticity. Well-being concerns what is good for the agent. If the pleasures that you enjoy are not *your* pleasures—they are derivatively inauthentic—why should we think that these contribute to your well-being or at least contribute to your well-being to the same extent that they would have had they been derivatively authentic? Suppose you don't care about whether your desires are authentic. Suppose that it is an engineered-in desire that moves you to action, but you are unaware that it is this sort of desire that is your effectively motivating desire. A good case can be made for the view that you are not deserving of, say, praise for the resulting action because the desire (and other elements) that drive you to action are not truly your own. Analogously, you may be indifferent to the authenticity of your springs of action, but if your pleasures are derivatively inauthentic—they are not truly your own—then, their not being authentic, it seems, should have some bearing on your well-being.

Again, the "standard" sort of retort ought to be anticipated: we shouldn't conflate what makes for a *good life* with what makes for a *prudentially good life*. If all your pleasures and displeasures are engineered into you in the absence of your knowledge of your being manipulated, your *life* may not be a good one although it may well be a life that is high in welfare value. But once again one may be fully aware of the ambiguity of the phrase "the good life," and may not be confusing the value of a *life* with the value of a *life for the one who lives it*. The claim is that authenticity has a bearing on *well-being*; it has a bearing on *prudential value* and not just on the value of a life. And the root intuition underlying this claim is simple enough: if a pleasure is not truly *your own*, then it is not as good (for you) as an otherwise similar pleasure that is truly your own (similar things are true about displeasures).

How do we resolve or break the following sort of dialectical stalemate that is familiar enough? On the one hand, there is the plausible intuition that if a factor bears on welfare, then its "presence" or "absence" must make a difference to the subject (truth, object-worthiness, or freedom, it is championed, are not factors of this sort). On the other hand, there is the competing intuition that things such as truth,

object-worthiness, and freedom may well affect welfare even if only as constituents of life atoms. (A sophisticated attitudinal theorist might, e.g., hold that freedom is intrinsically good, but its being intrinsically good makes no *independent* contribution to welfare; freedom contributes to *welfare*, it may be suggested, only as a constituent of life atoms that are core attributions of pleasure and displeasure.) Imagine a case in which new-wave neurosurgeons cause you to believe that, at a certain time, you are, at that time, taking intrinsic attitudinal pleasure in some object, when it is not in fact the case that you are doing anything of that sort. Concerning this object, you simply don't have the relevant attitude but because of the manipulation you don't know any better. Imagine, further, that you *never* discover that you have been duped. Presumably, we would not want to say that such mere belief contributes to your well-being. When it is personal well-being that is of interest, it matters whether you are really taking pleasure in the pertinent object. Imagine, next, that you *are* taking pleasure in the object, but only because you have been caused by the neurosurgeons to do so. Your pleasure is inauthentic. Assume that your attitudinal pleasure, in this case, is divorced from any sensory pleasure. If the authenticity or inauthenticity of your pleasure makes no difference to your well-being, why should the fact that your pertinent belief (in the prior case) is false make a difference to your well-being? One might respond that in the prior case, what is at issue is whether you do indeed take pleasure in the object. In the latter case, though, whether you take pleasure in the object is not in question; here the worry (if there is one) is with the authenticity of the pleasure. But this retort misses the relevance of juxtaposing the two cases. Reverting to the first case, imagine that your twin is just like you save that your twin really takes intrinsic attitudinal pleasure (attitudinal pleasure that is divorced from any sensory pleasure) in the pertinent object. Confining attention to the relevant life segments, your life feels to you from the "inside" just as your twin's life feels to her from the "inside." You both believe that you're taking pleasure in some object, but whereas your belief is false, your twin's belief is true. Whether your belief is true or not does not make a "difference" to you. Despite not making any such difference, regarding your personal well-being, it *matters* whether your belief is true or false, or so I am supposing any attitudinal hedonist should aver. Reconsidering the second case, imagine this time around that whereas your pleasure is inauthentic, your twin's pleasure in the same object is authentic. Again, given the relevant life segments (the segments in which both you and your twin take intrinsic attitudinal pleasure in the object), your life feels to you from the "inside" just as your twin's life feels to her from the "inside." Why, though, should such internal indiscernibility cast doubt on the view that the authenticity or inauthenticity of pleasures or displeasures can have an impact on personal well-being *if* indiscernibility in the prior case is irrelevant to whether mere belief that you are taking pleasure in something or true belief that you are doing so influences the truth of judgments of personal well-being?

The "make a difference to one's own life" intuition concerning factors that bear on personal well-being has a lot going for it. But it is not *quite* on target as this previous pair of cases suggest. None of this, of courses, *establishes* that factors such as truth, object-worthiness, or freedom are welfare-relevant factors, but at least one

barrier against taking such factors to be welfare-relevant is removed: it is false that, roughly speaking, if a factor does not "make a difference to one's own life," then that factor is not welfare-relevant. Considerations of authenticity try to capture the intuition that if, for instance, it is intrinsic attitudinal pleasures (and displeasures) that are the primary welfare-relevant factors, these things must be "truly one's own." Metaphorically speaking, a person may "have a self" that is authentic or inauthentic. Maybe your inauthentic self fares well in that it enjoys net intrinsic attitudinal pleasure (in a segment of its life). Why should this be pertinent to whether your authentic self fares well? Regarding prudential value, when we say that it is the agent's well-being that is at issue, we should be clear on who the agent is, and, again figuratively speaking, I am proposing that it is the "authentic agent" that is the proper subject. In short, we can say that considerations of authenticity seek to capture the intuition that welfare is *essentially personal*: it is the "authentic self" whose well-being is of paramount concern.

Chapter 6
Incompatibilism, Compatibilism, Desert, and Value

6.1 Subject's Desert-Adjusted Intrinsic Attitudinal Hedonism, Lives, and Worlds

Perhaps one of the most effectual ways to highlight the impact of freedom on various sorts of axiological appraisal is by exploring what incompatibilists and compatibilists should say about the influence of determinism on the intrinsic value of lives and worlds. In this chapter, I first list some source incompatibilist presuppositions. Then I advance a number of arguments to show that source incompatibilists should acknowledge the view that if our world is determined, and the lives of people contain at least some episodes of intrinsic attitudinal pleasure and some of intrinsic displeasure, then determinism can have a clear-cut impact on the value of lives (welfare value) and the value of worlds. Next, I propose that compatibilists should not be as perturbed as source incompatibilists about the influence of determinism on well-being and worldly value. Finally, I argue that at least one sort of libertarian, just like the source incompatibilist, has reason to be concerned about the possible effect of determinism on the intrinsic value of lives and worlds.

6.2 Some Source Incompatibilist Presuppositions

Source incompatibilists all accept the principle of ultimate origination (*Principle O*). Derk Pereboom, who is such an incompatibilist, advances this formulation of the principle:

> O: If an agent is morally responsible for deciding to perform an action, then the production of this mental action, a decision, must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control (Pereboom 2001, pp. 4, 43).[1]

Though I have some qualms about how, precisely, to interpret this principle (see Haji and Cuypers 2006, Haji forthcoming), its general thrust is fairly clear: if you

[1] Also see Kane (1996), esp. Chapter 5.

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

have no control of *any* sort over the provenance of your mental or other actions, then you are not responsible for those actions. Appealing to this principle, source incompatibilists argue that if determinism is true, all our behavior is, at the end, the upshot of the distant past and the laws of nature, things over which we have no control. Hence, it follows from *Principle O* and determinism that we have no responsibility-relevant control, direct or indirect, over anything we do. The source incompatibilist is committed to conceding that if determinism is true, not only are we not responsible for our decisions or overt actions, we are not responsible for any of our emotion-tokens, attitude-tokens, or feeling-tokens, regardless of their constitutional nature, owing to none of these tokens being (indirectly) free.

A second source incompatibilist presupposition is that if determinism is true, none of our decisions is derivatively authentic because the having of none of our springs of action, such as the having of desires and beliefs, is free. One may, though, be wary about this claim. It is true that if our world is determined, then the source incompatibilist should accept the view that no events are free and, hence, that the having of our desires and beliefs is not free. However, if our desires are not free, does it follow from their unfreedom that they are not "truly our own" or authentic? Should source incompatibilists not differentiate between unfree desires that are inauthentic and unfree desires that are "truly our own"? Whether this option is indeed available to *all* source incompatibilists may well be disputable. In any event, the source incompatibilists of interest endorse the following assumption that rules out this option for them:

> *No Difference*: There is no relevant and principled difference between an action whose causal history includes responsibility-undermining manipulation (as in the Ann/Beth scenario) and an action that has a more ordinary deterministic causal history, for, in either case, the actions issue from sources—the distant past plus the laws—over which agents have no control (see, for example, Kane 1996, 2000; Pereboom 2001, 2002 Chapter 4; 2002).

We said that actions that arise from inauthentic causal antecedents are derivatively inauthentic because their causal antecedents are inauthentic. If actions that derive from responsibility-undermining manipulation of the sort exemplified in the Ann/Beth case, as source incompatibilists acknowledge, are derivatively inauthentic, and if it is further assumed that a deterministic causal history is *not* relevantly different from a history involving apt manipulation, it seems that there are no grounds to distinguish between (the having of) causally determined desires that are authentic, in the relevant sense of "authentic," and (the having of) causally determined desires that are inauthentic.

As a third presupposition, it appears that source incompatibilists should accept PDA:

> *PDA*: If an agent receives a particular pleasure (or displeasure) upon performing a derivatively inauthentic act, A, which is an instance of pleasure of the same sort of pleasure that the agent would have deserved to receive in virtue of performing an act as "close" as possible to A with the exception of being derivatively authentic, then the particular pleasure (or displeasure) that this agent receives is not deserved.

Why so? Drawing, initially, on the Ann/Beth scenario, imagine that Beth's first posttransformation, derivatively inauthentic act is to give alms to the poor. It is highly plausible that Beth is not deserving of moral praise for this act. The reason is simple: the causal springs giving rise to the act are inauthentic. Some may well disagree with this verdict, but it does have its fair share of partisans who include compatibilists and both nonlibertarian incompatibilists who are source incompatibilists and libertarians.

In Section 4.7, we outlined the Beth/Beth* case. We imagined that for each act that (manipulation-free) Beth* performs in a segment of her life, (manipulated) Beth performs a type- or near type-identical act in a segment of her (Beth's) life. Assume that one of these sorts of act is giving alms to the poor. Beth*'s instance of this act is derivatively authentic, whereas Beth's instance of it is derivatively inauthentic. Both receive pleasure upon performing this act. Beth*'s pleasure is positive desert-based. She deserves this pleasure and she receives this pleasure. We asked whether the corresponding pleasure that Beth receives is *deserved*.

I ventured that the pleasure (or displeasure in an otherwise parallel case in which each receives displeasure instead of pleasure) is *not* deserved. I proposed that the pleasure is not deserved because the germane action, Beth's giving alms, causally issues from actional springs that are inauthentic and the action is, thus, derivatively inauthentic. A comparison may help. Beth is not morally praiseworthy for giving alms (or so, I assume, siding with the pertinent source incompatibilists on this score) because her action is derivatively inauthentic. Suppose someone heaps praise on Beth for what is perceived to be a praiseworthy deed—Beth's giving alms. This overt praise that Beth receives is not deserved. It is not deserved in virtue of Beth's not being praiseworthy for giving alms. Thus, the overt praise is not deserved because, ultimately, her giving alms is derivatively inauthentic. Similarly, the pleasure that Beth receives on giving alms is not deserved because, ultimately, her giving alms is derivatively inauthentic. If, then, one believes that one is not deserving of praise or blame for performing a derivatively inauthentic act (as source incompatibilists do), then one should believe, too, that one does not deserve pleasure (or displeasure) on the basis of performing such an act. Hence, source incompatibilists should accept PDA.

6.3 Some Principles

Before delving into arguments for the view that determinism can have an influence on well-being and on the value of worlds (assuming the source incompatibilist presuppositions exposed in the last section), it will be useful to list some principles, which I have advanced so far. They include the following.

(Pa): A pleasure (or displeasure) is (indirectly) free only if it derives from a decision that is (directly) free.
(Pb): A decision is (directly) free if and only if its agent has control in making it and it causally derives nondeviantly from authentic springs.

Thus,

(Pc): A pleasure (or displeasure) is (indirectly) free only if it issues from a decision which is such that the agent of the decision has control in making it, and this decision causally derives nondeviantly from authentic springs of action.

We said that an action-based pleasure (or displeasure) is a pleasure (or displeasure), which, if deserved, is deserved on the basis of an action-implicating desert base. So, somewhat loosely, an action-based pleasure is a pleasure that, if deserved, is deserved in virtue of an action being performed. Recall, a global induction case, such as the Ann/Beth case, suggests that if an action-based attitudinal pleasure (or displeasure) is deserved, then this pleasure (or displeasure) derives from authentic springs. Hence,

(DAIP): An action-based pleasure (or displeasure) is deserved only if it derives from authentic springs of action.

The arguments that follow in the next four sections appeal to one or more of these principles.

6.4 The Argument from Control

At the heart of the first argument is this thought: free intrinsic attitudinal pleasures are better than otherwise similar unfree pleasures. According to the incompatibilist, determinism undermines the freedom of our pleasures. Thus, determinism can affect well-being.

To develop the first argument more carefully, assume that our world is determined, incompatibilism is true (i.e., it is true that determinism is incompatible with free action and moral responsibility), our lives contain at least some episodes of intrinsic attitudinal pleasure and perhaps some of intrinsic displeasure, and that *favored symmetric freedom-sensitive hedonism* (the view, in brief, that free intrinsic pleasures are better than otherwise similar unfree pleasures and free intrinsic displeasures are not as bad as otherwise similar unfree displeasures) is the "life-ranking principle" of the correct life-ranking axiology. The first argument exploits principle ***Pa***. If the world is determined, no decisions are free and, hence, no pleasures or displeasures are free (***Pa*** tells us that pleasures and displeasures are free only if they derive from free decisions). This gives us the basis for the argument from control:

(1) If incompatibilism is true (and our world is determined), then no pleasures (or displeasures) are free.
(2) If no pleasures (or displeasures) are free, then this fact has a bearing on the intrinsic value of lives.
(3) Therefore, if incompatibilism is true (and our world is determined), then the fact that no pleasures (or displeasures) are free has a bearing on the intrinsic value of lives.

The rationales for the premises are straightforward. Line (1) is simply based on ***Pa***. Freedom-level control is the control an agent exercises in performing an action (mental or otherwise) that is required for moral responsibility. If incompatibilism is true and our world is determined, we lack freedom-level control in making any decision. Line (2) rests on the following sorts of consideration. Suppose that the lives of Koss and Floss contain otherwise type- or near type-identical episodes of intrinsic attitudinal pleasure save that all the pleasures in Koss's life are free (the "free life"), whereas all the pleasures in Floss's life are not free (the "unfree life"). The symmetric freedom-sensitive theory at issue yields the result that Koss's free life is better in itself for Koss than is Floss's unfree life for Floss. If we suppose that the two lives contain some intrinsic pleasures and some intrinsic displeasures, it is more complicated to determine the *favored symmetric* theory's ratings of these lives. Nonetheless, it will still be true that the freedom (or unfreedom) of the life atoms contained in these lives will have an impact on the value of these lives for their agents.

This sort of argument, with trivial changes, also sustains the verdict that if incompatibilism is true, then freedom has an impact on the intrinsic value of (determined) worlds, so long as it is assumed that freedom affects the value of world atoms.

6.5 The Argument from Desert

Fueling the second argument is the view that deserved intrinsic attitudinal pleasures are better than otherwise similar pleasures that are not deserved. According to source incompatibilists, determinism undermines desert of our action-based pleasures and displeasures. Thus, determinism can affect well-being.

The assumptions underpinning the argument from desert are these: our world is determined; incompatibilism is true; our lives contain at least some episodes of intrinsic attitudinal pleasure and some of intrinsic displeasure; subject's desert-adjusted intrinsic attitudinal hedonism (SDAIAH) is the correct *life*-ranking axiology; and the source incompatibilist's presuppositions, first, that no springs of action in a determined world are authentic and, second, that a manipulated causal history is not pertinently different from a deterministic causal history, are true. The argument from desert exploits principle DAIP. This principle says that action-based pleasures and displeasures are deserved only if they derive from authentic springs of action. If incompatibilism is true and our world is deterministic, then none of our action-based pleasures or displeasures is derivatively authentic, on the relevant source incompatibilist presuppositions. The argument is as follows:

(1) If incompatibilism is true (and our world is determined), then no pleasures (or displeasures) causally arise from a decision whose springs of action are authentic.
(2) If no pleasures (or displeasures) causally arise from a decision whose springs of action are authentic, then no action-based pleasures (or displeasures) are deserved.

(3) Therefore, if incompatibilism is true (and our world is determined), then no action-based pleasures (or displeasures) are deserved.
(4) If no action-based pleasures or displeasures are deserved, then this fact has a bearing on the intrinsic value of lives.
(5) Therefore, if incompatibilism is true (and our world is determined), then the fact that no action-based pleasures or displeasures are deserved has a bearing on the intrinsic value of lives.

Line (1) is based on the source incompatibilist's presupposition that no person's springs of action in a determined world are authentic. Line (2) draws on principle DAIP: an action-based pleasure or displeasure is deserved only if it derives from authentic springs of action. Line (3) is an intermediate conclusion. Line (4) looks to SDAIAH's ratings of otherwise similar free lives and unfree lives. To develop the rationale for (4), we start with a distinction as follows:

Application of Desert Bases to Worlds: *A desert base fails to apply in a world* if conditions at the world render it inappropriate to distribute goods on its basis. *A desert base applies in a world* if it is false that it fails to apply in this world.

If none of the potential recipients of desert in a world are capable of exerting conscientious effort, for example, then if exerting conscientious effort *is* a desert base or if a desert base implies that the recipient of desert is able to exert conscientious effort, this base fails to apply in this world even though other bases, such as innocent suffering, may well apply. Some worlds are worlds in which the only desert bases that apply are action-implicating; other worlds are worlds in which the only desert bases that apply are nonaction implicating; and still other worlds are ones in which both action-implicating and nonaction-implicating desert bases apply. Assume that Ernie and Bert live in worlds in which the only desert bases that apply are action-implicating desert bases. If Ernie and Bert deserve pleasures or displeasures, they deserve these things because of actions they perform. Imagine that each action-based intrinsic attitudinal pleasure that Ernie enjoys is a pleasure that he deserves to enjoy, and each action-based intrinsic displeasure that Ernie suffers is a displeasure that he deserves to suffer. In Ernie's "free life," there is a perfect "fit" between desert of pleasures (and displeasures) and receipt of pleasures (and displeasures). We introduce the notion of an unfree action-based pleasure.

Free action-based pleasures: An action-based pleasure is a pleasure, which, if deserved, is deserved because of some action. If this action is freely performed, then the action-based pleasure is free; if this action is not free, perhaps because it derives from inauthentic springs, then the action-based pleasure is not free.

Imagine that Bert receives action-based pleasures and displeasures as close to the ones that Ernie receives but that in his determined world, none of these pleasures and displeasures is free and so none is deserved. Assume, further, that each of Bert's action-based pleasures and displeasures is gratuitous. If SDAIAH offers an evaluation of *lives* by appeal to net amounts of attitudinal pleasure adjusted for subject's desert, Ernie's free life is better in itself for Ernie than Bert's unfree life for Bert.

To appreciate the import of this argument, in particular, to appreciate the fact that determinism may well negatively influence well-being (given certain source-incompatibilist presuppositions), I develop the Beth/Beth* case yet further.

Reflect, again, on the type- (or near type-identical) acts that Beth* and manipulated Beth perform at some time, *ts*, after Beth's recovery from transformation surgery at *t*. All of Beth*'s instances of these types of act are assumed to be derivatively authentic and positive desert-based; all of Beth's instances of these types of act are derivatively inauthentic. We said that for each episode of positive desert-based pleasure that Beth* receives because of an action she performs, Beth receives a corresponding pleasure that she does not deserve to receive. It is not that Beth *deserves* not to receive these pleasures; it is simply that these are not deserved—they are gratuitous. Each of Beth's corresponding pleasures is a neutral desert-based pleasure. Analogously, the corresponding displeasures that Beth receives are gratuitous; each is a neutral desert-based displeasure.

AXP1 implies that if a person gets the pleasure that she deserves, this is especially good, more so (as AXP3 implies) than if the person gratuitously receives the pleasure. For each positive desert-based pleasure that Beth* receives (in virtue of performing an action), there is a corresponding neutral desert-based pleasure that Beth receives. Each of these neutral desert-based pleasures that Beth receives is less good than the corresponding (positive desert-based) pleasure that Beth* receives.

Now for the pertinent displeasures, AXP2 implies that if a person gets the displeasure that she deserves, this is not so bad; it is less bad (as AXP3 implies) than if the person gratuitously receives the displeasure. For each negative desert-based displeasure that Beth* receives (because of performing an action), there is a corresponding neutral desert-based displeasure that Beth receives. Each of these neutral desert-based displeasures that Beth receives is worse than the corresponding (negative desert-based) displeasure that Beth* receives.

Assume that, with the exception of the set of pleasures and displeasures of concern—the positive desert-based pleasures that Beth* receives (because of actions she performs), the corresponding (neutral desert-based) pleasures that Beth receives, the negative desert-based displeasures that Beth* receives (because of actions she performs), and the corresponding (neutral desert-based) displeasures that Beth receives—Beth and Beth* are indiscernible with respect to all the pleasures and displeasures that contribute to their well-being. We can conclude that Beth*'s life is better in itself for Beth* than Beth's life is for Beth. For all of Beth's corresponding neutral desert-based pleasures are less good than Beth*'s positive desert-based pleasures, and all of Beth's corresponding neutral desert-based displeasures are worse than Beth*'s negative desert-based displeasures.

Allowing ourselves certain liberties and omitting certain details, we may give a simplified summary of this result in the following fashion. Imagine a certain life segment of Beth* and a "corresponding" life segment of Beth. Beth* receives intrinsic pleasures and displeasures in this segment; Beth receives corresponding pleasures and displeasures in her life segment. The life segments are indiscernible with respect to the raw amounts of pleasure and displeasure that they contain. (Recall, the "raw" amount of pleasure (displeasure) in an episode of attitudinal pleasure (displeasure)

is the product of the intensity (strength of attitude) and duration of that episode. This raw amount has not been adjusted, for example, for truth, or object-worthiness, or subject's desert.) Each of Beth*'s pleasures and displeasures that she receives is deserved because of actions that she performs; there is a perfect mesh between the pleasures and displeasures that she receives and the pleasures and displeasures that she deserves to receive. Each of Beth's corresponding pleasures and displeasures is gratuitous; these are not deserved. Other things being equal, deserved pleasures are better than pleasures that are not deserved; and deserved displeasures are less bad than displeasures that are not deserved. So this life segment of Beth* is intrinsically better for Beth* than is Beth's corresponding life segment for Beth.

The source incompatibilist presupposition, *No Difference*, says that there is no relevant difference between a deterministic causal history and a "manipulated causal history." With this assumption in mind, we may spin the tale of Beth and Beth* in a slightly different way. We introduce one significant change: assume that up until a certain time, *tc*, neither is the victim of manipulation and the two are indiscernible with respect to the pleasures and displeasures that each receives or fails to receive that contribute to their well-being. After *tc*, though, and up until a later time *tl*, although Beth does not fall victim to new-wave transformation surgeons, each of her deeds is causally determined. The rest of the story unfolds as before. Each (in the relevant life segment that extends from *tc* to *tl*) performs type- or near type-identical acts on the basis of which each receives intrinsic attitudinal pleasures and intrinsic displeasures. All of these acts of Beth* are authentic and all the pleasures (or displeasures) that she receives because of these acts are deserved. But all of Beth's corresponding acts are inauthentic, and the corresponding pleasures (or displeasures) that she receives because of them are not deserved; they are gratuitous. It follows from PDA and axiological principles AXP1–AXP3, that up until *tl*, Beth's life (the unfree life) is worse in itself for her than is Beth*'s life (the free life) for Beth*. Thus, given source incompatibilist presuppositions, determinism can detrimentally affect well-being.

Some people will reject line (4) of the argument from desert. They might claim that *Assumption Value* is false. That assumption, they might say, has the false implication that subject's desert can have an impact on *well-being* by having an impact on the values of episodes of intrinsic attitudinal pleasures and displeasures. Or, put in another way, these people will reject line (4) because they will insist that SDAIAH is the correct axiology for ranking *worlds* but not the correct axiology for ranking *lives*.

In the wake of this worry, the argument from desert can be modified to sustain the conclusion that determinism can have an impact on the value of worlds. Embellishing the Beth/Beth* case confirms that this is so. Imagine two worlds, each containing a single recipient of value. Beth* inhabits one of these worlds and Beth the other. Assume, first, that up until a certain time, *tc*, the values of the pleasures and displeasures that each receives adjusted for desert are identical. After *tc*, though, when each begins to perform actions, things change. Assume, second, that Beth's world is deterministic; hence, all of Beth's deeds are causally determined. But Beth*'s world accommodates free actions; all of Beth*s actions are free. Assume, third, that for each act that Beth performs after *tc* at some time, *ts*, in her world, Beth* performs

an act type- or near type-identical to this act in her world at *ts;* and for each act that Beth* performs after *tc* at some time, *tt*, in her world, there is a type- or near type-identical act that Beth performs after *tc* at *tt* in her world. Assume, fourth, that during her life segment that extends from *tc* to the time of her death, *td*, Beth* receives pleasures and displeasures, some of which she deserves to receive, others of which she does not deserve. Assume, fifth, that throughout her life, it is false that on any occasion Beth* deserves to receive displeasure but receives pleasure then instead; and it is false that on any occasion Beth* deserves to receive pleasure but receives displeasure then instead. During her life, there is a perfect fit between the pleasures (and displeasures) she receives and the pleasures (and displeasures) she deserves to receive. Of the positive desert-based pleasures that Beth* receives in her life, again the pleasures of concern are the pleasures Beth* deserves to receive in virtue of performing actions, such as giving alms to the poor. These are the "target" positive desert-based pleasures. (Similarly, we restrict attention to the target negative desert-based displeasures.) Assume, sixth, that for each (episode of) target positive desert-based pleasure that Beth* receives in her life, there is a corresponding pleasure that Beth receives, a pleasure that she receives upon performing an act type- or near type-identical to the one Beth* performs, this act of Beth*'s being the reason for Beth*'s receiving the apt episode of target positive desert-based pleasure that she receives. Given source incompatibilist presupposition *No Difference* (deterministic causal histories and manipulated causal histories are not relevantly different), each of Beth's deeds in the life segment extending from *tc* to *td* is derivatively inauthentic; we assume that each of Beth*'s corresponding deeds is derivatively authentic. It follows from PDA (if an agent performs an action that is inauthentic owing to this action's arising from causal antecedents that are inauthentic, and this action is such that, were it authentic, it would be an action in virtue of which the agent would deserve a pleasure or displeasure, then if the agent receives an episode of this sort of pleasure or displeasure upon performing the inauthentic action, the pleasure is not deserved) and SDAIAH that Beth's world is intrinsically worse than Beth*'s world. Thus, given source incompatibilist presuppositions, determinism can detrimentally affect the intrinsic value of worlds.

This conclusion invites the following three remarks. First, recall axiological principles AXP4 and AXP5 now adjusted so that they are pertinent to ranking worlds:

AXP4: Negative desert mitigates the intrinsic goodness of pleasure. If someone deserves displeasure but gets pleasure instead, the value of that pleasure for the world is mitigated.

AXP5: Positive desert aggravates the intrinsic badness of displeasure. If someone deserves pleasure but gets displeasure instead, the value of that displeasure for the world is aggravated.

One reason why we cannot draw the blanket conclusion that determinism casts a shadow over the value of all determined worlds (given the relevant source incompatibilist presuppositions) is because there are deterministic worlds in which a person deserves displeasure for certain reasons and gets pleasure instead, deserves pleasure

for certain reasons and gets displeasure instead, deserves pleasure for other reasons and gets pleasure, deserves displeasure for other reasons and gets displeasure, and gets gratuitous pleasure and displeasure. It is notoriously difficult to figure out the overall value of such worlds (even if we accept the pertinent source incompatibilist presuppositions).

I do not deny that (endorsing the source incompatibilist presuppositions) determinism may well *enhance* the value of some worlds. Imagine, for instance, that Flambo is a villain through and through but lives it up. The sole recipient of value in a world, there is a mismatch between the vast doses of action-based pleasure that he receives and the action-based pleasure that he deserves to receive. This injustice deflates the value of his world. In a one-value-recipient world as close as possible to Flambo's save for being deterministic, a counterpart of his, Flambo*, receives the same sorts (and doses) of intrinsic pleasure and displeasure that Flambo receives but each of these pleasures and displeasures is gratuitous. Given axiological principle AXP3, Flambo*'s world may be assumed to be better than Flambo's. Thus, determinism (and so hard incompatibilism) would preclude the intrinsic badness featured in Flambo's world. Clearly, then, it is possible that hard incompatibilism would not have an adverse effect on the value of a deterministic world if there were a balance between the intrinsic badness lost and intrinsic goodness lost, or if there were more intrinsic badness lost than intrinsic goodness lost, in that world. But there are worlds and there are worlds: hard incompatibilism will have a detrimental impact on the value of some but not on the value of others. Its effect, positive or negative, will turn on *contingent* features of the world such as the distribution of pleasures and displeasures among its recipients of value. Hence, the tempered conclusion to which we are entitled is that it is possible that determinism affects (either positively or negatively) the intrinsic value of worlds (on source incompatibilist presuppositions) as a result of having an impact on the desert of pleasure or displeasure.

Second, an obvious maneuver to attempt to circumvent even this cautious conclusion is to deny one or more of the source incompatibilist presuppositions. So, for instance, one may well be skeptical about assumption *No Difference* or the principle of ultimate origination. Another tactic would be to cast doubt on SDAIAH. My interest, however, is in getting clear on what stance source incompatibilists who accept these presuppositions, and who accept this axiology, should adopt concerning determinism and the value of worlds.

Finally, one might well wonder about the import of the tempered conclusion. An interesting theme in the recent literature on free will is that the value of having the freedom that moral responsibility requires has been overestimated. Prominent source incompatibilists have proposed that even if we lack this freedom, as we would if our world were deterministic, our lives would not be so detrimentally affected (Waller 1990, Honderich 1993, 2002; Pereboom 2001, Double 2004). Pereboom, for instance, has argued that though lack of this freedom undermines appraisals of moral praiseworthiness and moral blameworthiness, our being without this freedom leaves intact other sorts of moral appraisal (such as appraisals of moral obligation, right, and wrong) is no bar to developing an ac-

ceptable position on managing criminal behavior and on moral education and reform, and fails to affect significantly interpersonal relationships and "life-hopes" (Pereboom 1995, 2001, 2002).[2] The source incompatibilist presuppositions to which Pereboom subscribes, though, muddies this comforting picture considerably if it is true, assuming SDAIAH offers an evaluation of lives, that determinism can adversely affect well-being. Even if this assumption is false, the rosy picture is tainted if determinism can adversely affect the value of worlds. If our world is a hard incompatibilist one that lacks the freedom that responsibility requires, for all we know, rather than being one of the worlds in which worldly value is untouched or enhanced by hard incompatibilism, it may be one in which worldly value is detrimentally affected by hard incompatibilism.

6.6 The Argument from Authenticity

Before we leave the argument from desert, note that an argument similar to this desert argument, in that it also invokes considerations of authenticity but that relies on *favored symmetric freedom-sensitive hedonism* (and not on SDAIAH) as the *life*-ranking axiology, supports the same conclusion as the conclusion of the second desert argument. The assumptions underlying this third argument are these: our world is determined; incompatibilism is true; our lives contain at least some episodes of intrinsic attitudinal pleasure and perhaps some of intrinsic displeasure; favored symmetric freedom-sensitive hedonism is the correct life-ranking axiology; and the source incompatibilist presupposition that all the springs of our actions in a determined world are inauthentic is correct. This argument draws on principle ***Pc***: pleasures and displeasures are free only if they derive from authentic springs of action. It can be formulated in this way:

(1) If incompatibilism is true (and our world is determined), then no pleasures (or displeasures) derive from authentic springs of action.
(2) If no pleasures (or displeasures) derive from authentic springs of action, then no pleasures or displeasures are free.
(3) If no pleasures (or displeasures) are free, then this fact has a bearing on the intrinsic value of lives.
(4) Therefore, if incompatibilism is true (and our world is determined), then the fact that no pleasures (or displeasures) are free has a bearing on the intrinsic value of lives.

Line (1) derives from the source incompatibilist presupposition that no springs of action in a determined world are authentic. Line (2) rests on ***Pc***. Line (3) of this argument is identical to point (2) of the argument from control and, thus, it has the same rationale as that line.

---

[2] For a view similar to the view that some important "life-hopes" must be abandoned if determinism is true but that other life-hopes that matter to us can be retained, see Honderich (1988, 1993 and 2002).

6.7 Compatibilism, Well-Being, and the Value of Worlds

Compatibilists affirm that determinism is compatible with free action and moral responsibility. Compatibilism is true if and only if some compatibilist account of free action and responsibility is true. Suppose that compatibilism is true and our world is deterministic. What should compatibilists say about the effect of determinism on well-being or on the value of worlds? Reconsider the three arguments discussed in the prior three sections, this time recast as arguments designed to show that compatibilism threatens well-being or the intrinsic value of worlds. Understood in this fashion, are these arguments sound? To formulate the first revised argument that modifies the argument from control, we assume that free pleasures are better than otherwise similar unfree pleasures, free displeasures are not as bad as otherwise similar unfree displeasures, and that (given principle ***Pa***) if a pleasure (or displeasure) is free, it derives from a decision that is free. The argument can be presented in this way:

Argument 1

(1) If compatibilism is true (and our world is determined), then no pleasures (or displeasures) are free.
(2) If no pleasures (or displeasures) are free, then this fact has a bearing on the intrinsic value of lives.
(3) Therefore, if compatibilism is true (and our world is determined), then the fact that no pleasures (or displeasures) are free has a bearing on the intrinsic value of lives.

A compatibilist should claim that line (1) is false because decisions can be directly free, and thus, pleasures or displeasures can be indirectly free even if our world is determined. But a compatibilist should also grant that there are many situations in life in which our decisions are not free. Pleasures and displeasures that issue from such decisions will not be free either.

The next argument is the suitably amended argument from desert. We may derive this argument simply by replacing each occurrence of "incompatibilism" in the argument from desert with "compatibilism." It suffices to reproduce the first premise of this derived argument. It says

(1x) If compatibilism is true (and our world is determined), then no pleasures (or displeasures) causally arise from a decision whose springs of action are authentic.

Deriving the third modified argument in a similar fashion, its first premise affirms the following:

(1y) If compatibilism is true (and our world is determined), then no pleasures (or displeasures) derive from authentic springs of action.

Clearly, the compatibilist should eschew both premises (1x) and (1y). Compatibilists do not claim that determinism precludes our springs of action from being authentic. They should admit, though, that even in a determined world, actional springs may well be inauthentic under various circumstances. Many compatibilists,

for example, accept the verdict that manipulated Beth's engineered-in pro-attitudes, values, beliefs, and so forth are inauthentic and that Beth is not responsible for decisions that express these inauthentic springs.

In summary, just as compatibilists deny that determinism poses a threat to the freedom that moral responsibility requires, so compatibilists should deny that determinism per se poses a threat to personal welfare or to the value of worlds by threatening either the freedom of pleasures or displeasures or the desert of pleasures or displeasures.

6.8 Libertarianism, Well-Being, and the Value of Worlds

One of the source incompatibilist presuppositions is that there is no pertinent difference between a causal history with manipulation and a causal history that is deterministic but free of manipulation. This assumption plays a vital role in, for example, the original argument from desert. Just like many compatibilists, there are libertarians, though, who reject the view that there is no germane difference between manipulated causal histories and deterministic causal histories. Since such libertarians insist that determinism undermines the freedom of our decisions, attitudes, choices, and so forth, like source incompatibilists these libertarians should judge that determinism may well have a detrimental influence on well-being or on the value of worlds. But what of worlds in which, for instance, our actions and attitudes have *indeterministic* event-causal histories? Owing to a plethora of libertarian views, drawing sweeping generalizations regarding what stance libertarians who insist that free actions be indeterministically caused should take on the consistency of freedom and axiological judgments having to do with welfare and worldly value would be imprudent. We can, though, say something about what position so-called "modest libertarians" should adopt on the impact of indeterminism on well-being or on the value of worlds.

A modest event-causal libertarian account, unlike a noncausalist one, postulates that the freedom or control that free action and moral responsibility require is a variety of causal control.[3] The account shares central features with the most promising compatibilist views on free action. It demands that to choose or act freely, an agent must be able to engage in practical reasoning and to guide her behavior in light of the reasons that she has. In addition, an event-causal account makes no appeal to metaphysically exotic agents, such as Cartesian minds or to agent-causation, in its explication of the kind of control acting freely requires. Further, like its compatibilist competitors, it dictates that the actions (mental or otherwise) for which an agent is responsible be the outcome of causal processes. A free decision or overt action is one that is made for reasons and its being made for reasons consists, partially, in its being caused in an appropriate and nondeviant way by the agent's having those reasons.

[3] Noncausalist accounts are developed and defended by Goetz (1988); Ginet (1990, 1997); and McCann (1998).

Such a modest libertarian account differs from a compatibilist one in that it insists that, at some point or points in the causal trajectory to action, the causation among actional events (such as the having of a desire and the forming of a best judgment concerning what one should do) be indeterministic. This requirement answers to the libertarian's demand that free actions are not causally determined.

Sketching a distinction between active and ultimate control will be useful. Active control is a species of causal control. Such control concerns the direct causal production of agent-involving events, such as the agent's having certain values, desires, and beliefs, his making a certain evaluative judgment, his forming a certain intention or arriving at a certain decision, his executing an intention, and his performing a nonmental action. Ultimate control is the sort of control that one has, for instance, in making a decision provided that one is the "ultimate originator" of one's decision. Modest libertarians characterize ultimate control in terms of the absence of deterministic causation. The rough idea is that, if an agent's decision is indeterministically caused by prior agent-involving events, such as the agent's having of reasons, then it is false that there are conditions "external" to the agent (such as the obtaining of the natural laws and the nonrelational facts of the world at some past time) that are minimally causally sufficient for the agent's making that decision.[4] In this way, the agent is the "final source" of her decisions.

I have argued (in other works) that the active control that modest libertarianism secures is no different than the active control its best compatibilist rival secures (Haji 2004a; see, also, Clarke 2003, p. 27). I have also argued (in other works) that the variety of ultimate control that modest libertarianism supplies—something that no compatibilist competitor can provide—does not contribute to free action (see, e.g., Haji 2004a, 2005a). Thus, regarding the *control requirements* of free action, modest libertarianism *does not differ* from its best compatibilist adversary.[5] It follows that if the control that compatibilist theories provide is deemed inadequate for free action and moral responsibility, modest libertarian theories should be regarded as inadequate in this respect as well. But then the modest libertarian should claim that intrinsic attitudinal pleasures and intrinsic displeasures in indeterministic worlds (just as in deterministic worlds) are *not* free. If these attitudes are not free in these worlds, such libertarians are committed to the result that even indeterminism can have a pronounced effect on well-being or on the value of worlds, as a suitably adjusted version of the argument from control confirms.

---

[4] See, e.g., Mele (1995, p. 211) and Kane (1996, p. 35). A more detailed explanation of the distinction between active and ultimate control may be found in Haji (2006); forthcoming, Chapter 2.

[5] There is a complication, here, that I sidestep. There is a species of control, "antecedent actional control," that I believe responsibility requires and that is not supplied by modest libertarian views.

Chapter 7
Freedom, Obligation, and the Good

7.1 Obligation, Freedom, and the Value of Worlds

If there is a necessary connection between obligation (or the *right*) and the *good*, pleasure is the good, and freedom has an impact on the good, then it would seem that freedom has an impact on obligation. Further, if determinism undermines freedom, as incompatibilists claim, then determinism, too, it would seem, should have an influence on obligation. Whether determinism does indeed affect obligation is the broad topic of this chapter.

Our principal concern with obligation in what follows is not with the concept of moral obligation; rather, it is with a normative account that specifies necessary and sufficient conditions for something's being (overall) morally obligatory. With respect to this account, it has seemed obvious to many (though not to many others) that our fundamental moral obligation is to do the best we can—to make the world as good as we can. Roughly, on this view, as of some time, a person ought to bring about the states of affairs that occur in the intrinsically best worlds then accessible to him or her (see, e.g., Moore [1903] 1962, Feldman 1986, Zimmerman 1996). According to this normative theory, the obligatory act is part of the best life available to the agent; intuitively what you ought to do as of a time is what you do in the best life then open to you (Feldman 1997, p. 18). More rigidly, if a possible world, w, is accessible to agent, s, at time, t, and no better world is also accessible to s at t, then w is best for s at t. The basic principle of obligation can now be stated in the following way:

Best: As of t, s ought to see to the occurrence of state of affairs, p, if and only if p is true at all the bests for s at t and p is not true at some worlds accessible to s at t.

For our purposes, many fine details of this principle can safely be set aside. We simply record that it is this account of moral obligation that is presupposed in what follows.

This sort of view about obligation requires a ranking of possible worlds in terms of intrinsic value. Another assumption of this chapter is that some version of attitudinal hedonism is the axiology for such a normative theory. It is highly credible that the value of a world is at least partly a function of how just the world is. So I start with the view that subject's desert-adjusted intrinsic attitudinal hedo-

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

nism (SDAIAH) is the pertinent axiology. I argue that this axiology faces an initial problem: it generates rankings of certain worlds that are suspect. I propose that if SDAIAH is modified so that the values of its atoms are adjusted to accommodate the influence of both desert and freedom, the resulting freedom-sensitive axiology overcomes this problem. I advance competing accounts of world-ranking attitudinal hedonism that are both freedom- and desert-sensitive, and I examine what effect, if any, determinism has on obligation, given each of these accounts.

I then turn to another objection, which, if sound, undermines even freedom-sensitive variations of SDAIAH. SDAIAH is a version of *totalism*. *Totalism* is the view that the value of a world is the sum of the values its denizens receive. The intuitive appeal of totalism is evident: it has seemed apparent to many that a world's value cannot differ from the sum of the values of those who receive value in the world; the value of a world must be distributed among its recipients of value (see, e.g., Mill [1861] 1957, Chisholsm 1986). However, Derek Parfit has argued that all totalistic axiologies succumb to the problem of the *repugnant conclusion* (1984, p. 383). The problem, as we said, can be summarized in this way: totalist views rate highly populous worlds in which each recipient of value receives, over the course of her entire life, a nominal amount of some primary intrinsic good, say, net attitudinal pleasure, as better than a less populous world in which each recipient of value fares much better than each such recipient of the highly populous world. But because it contains relatively fewer recipients of value, the worldly value of the less populous world is lower than that of the highly populous world. Many find this rating repugnant. I argue that the repugnant conclusion does not refute totalism.

Finally, if we pay attention to the connection between desert and authenticity—in particular, if we keep in mind the principle that an action-based pleasure (or displeasure) is deserved only if it derives from authentic springs of action—the freedom-sensitive version of SDAIAH that I favor reveals a connection between moral obligation, on the one hand, and the inauthenticity of our springs of action, on the other: it is possible that a person not be morally obligated to perform an action that causally issues from inauthentic springs of action *owing to* the inauthenticity of these springs.

7.2 Determinism, Alternative Possibilities, and Obligation

One link between freedom and obligation is that obligation requires alternative possibilities. To explain, the account of moral obligation according to which we ought to do the best we can has as theorems the "ought" implies "can" principle (principle **K**) and the principle that it is obligatory to do something just in case it is wrong not to do it (principle **OW**):

K: If it is overall (and not just prima facie) obligatory for one to do something, then one can do it, and if it is (overall) obligatory for one to refrain from doing something, then one can refrain from doing it.

OW: It is (overall) obligatory for one to do something if and only if it is (overall) wrong for one to refrain from doing it.

If we accept **K** and **OW**, we may confirm that there is a requirement of alternative possibilities for the truth of judgments of moral obligation, right, and wrong (or, in brief, for judgments of obligation). First, note that **K** and **OW** entail that an agent can do moral wrong only if the agent had an alternative:

WAP: If it is wrong for one to do something, then one can refrain from doing it.

Barring cogent reasons to believe otherwise, if we assume that "ought" implies "can," there is little reason not to assume, too, that "wrong" (and "right") imply "can." For the freedom- or control-relevant presuppositions of obligation, it would seem, should also be the very ones of wrong and right. If we grant that "wrong" implies "can," we can show that obligation requires alternative possibilities:

(1*) If it is obligatory for one to refrain from doing *a*, then it is wrong for one to do *a* (from **OW**).
(2*) If it is wrong for one to do *a*, then one can do *a* (from the "wrong" implies "can" analogue of **K**).
(3*) Therefore, if it is obligatory for one to refrain from doing *a*, then one can do *a*.

There is no similar way to derive the proposition that rightness, likewise, requires alternatives. This is because even if it is agreed that "right" implies "can," there is no principle like **OW** that will allow us to infer that "right" implies "can refrain." Nevertheless, it is very plausible that "right" does imply "can refrain." For suppose an agent, Jones, is in a situation in which Jones cannot refrain from pushing an innocent child off the pier to her death. Perhaps Jones is in a "Frankfurt-type" situation in which Jones supposedly freely pushes the child off the pier but could not have refrained from doing so.[1] It is neither wrong nor obligatory for Jones to perform this act in his circumstances because there is a requirement of alternative possibilities for wrongness and obligation. It seems, then, that there are two principal options. An act has a *primary morally deontic status* only if it is morally right, wrong, or obligatory. Either pushing the child off the pier is permissible for Jones or this act lacks a primary morally deontic property altogether, not being morally right, wrong, or obligatory; it is, we can say, "amoral" for Jones. The latter is more plausible than the former. After all, Jones' act is the cold-blooded killing of an innocent child; for that matter, his act could have been any other heinous deed. In Jones' circumstances, if pushing the child to her death were not amoral for Jones, this act would be permissible—that is, *morally right*—for Jones, a result that resists acceptance. Insisting, again, that in the absence of persuasive reason to think otherwise, the freedom requirements of "right" parallel those of "wrong" and "obligation," enables us to circumvent this unpalatable result. Hence, "right," too, implies "can refrain."

[1] On Frankfurt-type examples, see (Frankfurt 1969).

We conclude that there is a requirement of alternative possibilities for obligation: no one can perform an act that is obligatory unless one could have done otherwise. If determinism is true, though, then arguably we never have alternatives. Hence, determinism is incompatible with obligation. Since I have developed and defended this line of reasoning elsewhere (Haji 2002), I shall not here pursue it any further. Instead, I wish to explore a second, less direct link between freedom and obligation. I propose that the correct world-ranking axiology should be freedom-sensitive. If it is, then it would seem that determinism may affect obligation as a result of affecting worldly value.

7.3 Ross's Objection Revisited

Feldman, as we have remarked, develops SDAIAH as the correct world-ranking axiology (partly) in response to Ross's *two worlds objection*. Assume that this objection has as its quarry the simple theory suitably modified to apply to worlds. To rehearse the objection, Ross imagines two worlds qualitatively identical with respect to what we stipulate is the intrinsic attitudinal pleasure and intrinsic displeasure contained in the lives of people in these worlds, but different in that, in one world, those who receive pleasure are virtuous and those who receive displeasure are vicious whereas, in the other, those who receive pleasure are vicious and those who receive displeasure are virtuous. Ross's objection is that the simple theory rates the worlds as equally valuable, whereas he judges the former to be better than the latter. I agree with Ross's evaluation of these worlds as does Feldman.

Since SDAIAH has atoms whose values are adjusted to reflect the extent to which subjects deserve or do not deserve the intrinsic attitudinal pleasures and intrinsic displeasures that they receive, Feldman concludes that this theory yields the rankings of the two worlds that Ross prefers. Despite this being so, there is still a remaining concern with some of SDAIAH's rankings of worlds.

7.4 A Remaining Problem Concerning Freedom with *Subject's Desert-Adjusted Intrinsic Attitudinal Hedonism*

Imagine that an entire world, the Matrix, is a huge, slimy womb in which its occupants are nourished. Every person in this nutrient-rich world begins life as a zygote. Their brains are connected to a sophisticated computer. As they mature into individuals who can adopt or assume attitudes, the computer stimulates suitable centers in their brains with the result that each is made to take intrinsic attitudinal pleasure in various objects; none takes intrinsic displeasure in anything. In this way, each occupant of the Matrix lives a make-believe life. Not one of these individuals is aware that he or she is living in the Matrix and is being made to take pleasure in different things. Intuitively, the worldly value of the Matrix leaves a lot to be desired.

Maybe some people will recommend that this sort of example calls into question all world-ranking forms of attitudinal hedonism. On my view, though, the example

motivates revising SDAIAH. The thought is that just as justice can affect the intrinsic value of a world, so, too, can freedom. Freedom should, hence, be a constituent of the atoms that contribute to worldly value.

Ponder what each denizen of the Matrix deserves. We have noted that there are many different desert bases including excessive or deficient past receipt of certain goods, moral worthiness, legitimate claims, the amount of time and effort expended in acquiring various goods, and so on. It is not so far fetched to suppose that no denizen of the Matrix has a positive desert base—no person in this world deserves any pleasure; no one has worked hard, no one is morally worthy; no one has been deprived of past pleasure, and so on.

Similarly, it is not so far fetched to suppose that no one in the Matrix has a negative desert base—no person in this world deserves any displeasure. Maybe each of the inhabitants of the slimy womb has a neutral desert base—he or she deserves neither pleasure nor displeasure. Neutral desert base neither enhances nor mitigates the intrinsic goodness of pleasure; nor does neutral desert base enhance or mitigate the evil of displeasure. The intrinsic value of any such episode of intrinsic pleasure (or intrinsic displeasure) is directly proportional to the amount of pleasure (or displeasure) it contains.[2] It appears that an intrinsic attitudinal hedonism that adjusts the intrinsic value of episodes of intrinsic attitudinal pleasure and displeasure for subjects' desert generates the result that the Matrix is not such a bad world after all; indeed, it may be a highly good world: each occupant of that world has a neutral desert base; each takes intrinsic attitudinal pleasure in many and various objects and no displeasure in any object. But it seems that this world, in which each recipient of value is captive, is just as bad or worse (perhaps even far so) than Ross's unjust world.

Maybe some will rejoin that a 20-year old in the Matrix has been deprived of freedom. Specifically, they might claim that although this person enjoys many intrinsic attitudinal pleasures, none of these pleasures is free. In virtue of his deficient past receipt of free pleasures, he deserves the pleasures of having or experiencing *free* pleasures. (Compare this with a case in which, in virtue of a person's deficient past receipt of good meals, the person deserves the pleasures of a good meal.) So it is false that no person in the Matrix has positive desert.

The objection, though, has shortcomings. The most glaring is that it is controversial whether being deprived of free attitudinal pleasures is a legitimate desert base. This is contentious, in part, because it is difficult (to say the least) to isolate necessary and sufficient conditions for something's being a desert base. Consider, for instance, the proposal that desert presupposes moral responsibility:

DB1: If agent *s* deserves something in virtue of *y*, then *s* is morally responsible for *y*.[3]

---

[2] See Section 7.5 for a complication having to do with whether or not the pleasures or displeasures are free.

[3] For discussion on the connection between desert and responsibility, see, for example, Rachels (1978), Sadurski (1985), Feldman (1995, 1996), Cupit (1996) and McLeod (2003).

DB1, though, is suspect. Babies, in virtue of, among other things, being (little) persons deserve respect but they are not responsible for being persons. Even if DB1 were in the right ball park, it would not imply that deficiency of free attitudinal pleasures, given the predicament of the persons in the Matrix, is a desert base. For not one of these persons is morally responsible for his or her deficiency.

Another view links desert bases with reactive attitudes such as admiration, resentment, gratitude, and indignation. The pertinent principle is something of this sort:

DB2: *x* is a desert base if and only if *x* is the basis of a reactive attitude.[4]

A primary problem with DB2 is that it remains moot what comprises "the basis" of the reactive attitudes. A popular view is that the reactive atttitudes presuppose moral responsibility.[5] If you feel indignation because somebody has taken unfair advantage of you, the indignation, if "appropriate," presupposes that the person is morally responsible for taking advantage of you. If you admire someone because of something she does, the admiration, again if "appropriate," presupposes that the person is morally responsible for that deed. If responsibility, though, is the sole basis of the reactive attitudes, then DB2 implies that there is only one desert base. This seems false. You may feel compassion because the child has been born with a serious deformity. The child, though, is not morally responsible for being born or for its deformity. Needless to say, there are contenders other than DB1 and DB2. My modest claim is simply that it is far from obvious whether deficiency of enjoying free attitudinal pleasures is a legitimate desert base.

A second problem with the rejoinder is that the 20-year old takes no intrinsic displeasure in anything. So even if we grant that he has a positive desert base in virtue of not having enjoyed free attitudinal pleasures, his being deprived of these free pleasures fails to aggravate the evil of any displeasure that he experiences owing to the fact that he experiences no displeasure. Still, on the assumption that he deserves pleasures that he fails to get, it is intrinsically bad for him to get nothing instead. However, since he enjoys vast doses of intrinsic attitudinal pleasure that he receives gratuitously—he does not deserve them—and since his situation is not relevantly different from anyone else's in the Matrix, it appears that SDAIH still generates the verdict that the Matrix is a good world.

Another objection is that the 20-year old has a positive desert base simply in virtue of being a person. Even granting this, the desired moral of the example is not disparaged. Insofar as the 20-year old deserves various pleasures because he is a person and receives them, SDAIAH implies that receipt of these pleasures *augments* the worldly value of the Matrix. This, if anything, provides even greater impetus to amend SDAIAH in a fashion in which the resulting theory yields a more intuitively acceptable rating of this world. The rating should imply that the Matrix is not such

---

[4] See, for example, Miller (1976). For further discussion on this view, see, for instance, McLeod (2003).

[5] See, for example, Strawson (1962) and Fischer and Ravizza (1998).

a good world, its deficiency in value being attributable to the unfreedom of the pertinent attitudes of its recipients of value.

We have said that an episode (or atom) of attitudinal pleasure that you enjoy is free if you freely take pleasure in some object. Just as a desire may be engineered into you against your will, so you may be made to adopt or assume certain attitudes against your will. Such engineered-in attitudes would not be free. "Free" in the last sentence has the same sense as the sense it has in utterances such as, "since all other conditions of responsibility were satisfied and her action was free, she was morally blameworthy for her deed."

It appears that SDAIAH yields an incorrect rating of the Matrix because this theory is insufficiently sensitive to whether episodes of intrinsic attitudinal pleasure and intrinsic displeasure are free. If we assume that takings of pleasure diminish the value of a world if they are unfree takings of pleasure (even though they may be takings of pleasure in true states of affairs) or altogether render such pleasures intrinsically worthless, then the attitudinal hedonist can account for the intuitive judgment that the worldly value of the Matrix is far lower than what SDAIAH makes it out to be.[6]

7.5 Varieties of Freedom-Sensitive World-Ranking Axiologies

In Chapter 3, we introduced different versions of freedom-sensitive life-ranking axiologies. Each version is a form of attitudinal hedonism. Corresponding to each of these life-ranking theories is a world-ranking theory. Let's formulate one of these world-ranking theories, the robust variety. On this theory, the world atoms all look relevantly like the following:

World Atom-1: At noon on Tuesday, October 16, 2001, Bob freely takes intrinsic attitudinal pleasure of intensity +8 in the fact that Bob's beer is frosty cold.

There are corresponding atoms or basics concerning intrinsic displeasure. According to this theory, only free episodes of intrinsic attitudinal pleasures and intrinsic displeasures have intrinsic value; unfree pleasures and displeasures are intrinsically worthless. The theory may be formulated in the following fashion:

Robust Freedom Sensitive Intrinsic Attitudinal Hedonism

i. Every free episode of intrinsic attitudinal pleasure is intrinsically good; every free episode of intrinsic displeasure is intrinsically bad.
ii. The intrinsic value of a free episode of intrinsic attitudinal pleasure is equal to the amount of pleasure contained in that episode; the intrinsic value of a free episode of intrinsic displeasure is equal to—(the amount of displeasure contained in that episode).
iii. The intrinsic value of a world is entirely determined by the intrinsic values of the free episodes of intrinsic attitudinal pleasure and displeasure contained in the world, in such

[6] Some pluralists may propose that (un)freedom and (dis)pleasure are independent sources of intrinsic value. The interest here, though, is both with monistic hedonism and a view about worldly value.

> a way that one world is intrinsically better than another if and only if the net amount of (free) intrinsic attitudinal pleasure in the one is greater than the net amount of (free) intrinsic attitudinal pleasure in the other.

Again, the theory may be enriched with other considerations thought pertinent to the intrinsic value of worlds. One might want, for instance, to adjust the values of pleasures and displeasures for freedom and subjects' desert. Indeed, I think that it is highly plausible that the correct world-ranking axiology should take into account both subject's desert and freedom. However, since our concern is first and foremost with freedom, I prefer to keep things simple and work initially with pristine freedom-sensitive varieties.

The robust theory yields the result that the Matrix is not intrinsically good. Suppose, though, that everyone in the Matrix were made to take intrinsic displeasure in various states of affairs. Many would be inclined to judge that such a world ("Matrix Evil") is an evil world. The robust theory, though, implies that Matrix Evil is intrinsically worthless. As some will see it, another concern with the robust theory is that freedom has a less drastic effect on the value of worlds (really, on the value of world atoms) than what the theory implies. The intrinsic value of a positive world atom is enhanced if the atom is free, and perhaps a negative atom is either worse if it is free or less bad if it is free. So let us, then, consider more modest versions of world-ranking attitudinal hedonistic axiologies.

On the asymmetric version, freedom enhances the value of a positive world atom—it enhances the values of episodes of intrinsic attitudinal pleasures—but has no effect on the value of negative atoms—it has no effect on the values of intrinsic displeasures. The theory implies that the intrinsic value of a world is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and intrinsic displeasure contained in that world, in such a way that one world is intrinsically better than another if and only if the net freedom-adjusted amount of intrinsic attitudinal pleasure in the one is greater than the net freedom-adjusted amount of that sort of pleasure in the other.

The theory does yield the desired outcomes that the Matrix in which each person is made to take intrinsic pleasure in various objects but suffers no intrinsic displeasure is not intrinsically good, and that Matrix Evil in which each person is made to take intrinsic displeasure in various objects but enjoys no intrinsic pleasure is intrinsically bad. Still, there seems to be little reason to deny that freedom (or lack of it) should affect the values of intrinsic displeasures if freedom affects the values of intrinsic pleasures. This may impel us toward the symmetric varieties.

On the first symmetric variety, free intrinsic pleasures are better than otherwise similar unfree pleasures, and free intrinsic displeasures are worse than otherwise similar unfree displeasures. On the second variety, again, free positive world atoms are better than otherwise similar unfree atoms, and free negative world atoms are better than (i.e., they are not as bad as) otherwise similar unfree atoms. Again, for reasons that I previously explained (see Section 3.3.3), I am drawn more to the second symmetric theory. I propose that a well worked out world-ranking axiology should attempt to partner this second symmetric version with considerations of

subject's desert. The resulting theory would be a cross between SDAIAH and this second version. It will be useful to outline the contours of such a theory.

On this hybrid theory, the adjustment for desert is to be made in accordance with principles AXP1—AXP5; the adjustment for freedom is to be governed by this principle:

> *Adjustment for Freedom*: Free pleasures are better than otherwise similar unfree pleasures, and free displeasures are not as bad as otherwise similar unfree displeasures.

So, for instance, if a person deserves an episode of pleasure and receives it, the exact match between receipt of pleasure and desert of pleasure enhances the value of this episode (for the world). If this episode is free, its value (for the world) is enhanced still further in virtue of its being free. Or if a person receives an episode of displeasure that she deserves, the favorable fit between receipt and desert makes the displeasure less bad. If the displeasure is free, its negative value is attenuated even further.

Let the values of the episodes of intrinsic pleasures and displeasures that a subject experiences be adjusted for desert in accordance with principles AXP1—AXP5 and for freedom in accordance with *Adjustment for Freedom*. We assume that there is a measure that accurately reflects the dual influence of desert and freedom on the values of positive and negative world atoms. Refer to this new measure of value as "subject's freedom- and desert-adjusted intrinsic value." Freedom-sensitive SDAIAH is the view that the value of a world is the sum of the subject's freedom- and desert-adjusted values of the intrinsic attitudinal pleasures enjoyed and intrinsic displeasures suffered in that world.[7]

Freedom-sensitive SDAIAH implies that the Matrix is overall good, but it is far worse than an otherwise similar world in which its denizens lead "real lives." And the theory implies that Matrix Evil is worse than an otherwise similar world in which its inhabitants lived "real lives."

Finally, *pain adjusted freedom-sensitive intrinsic attitudinal hedonism* adjusts the intrinsic values of episodes of intrinsic displeasure to reflect the extent to which these episodes are free and it pronounces, as intrinsically worthless, episodes of intrinsic attitudinal pleasure that are unfree. On the first variation of this theory, unfree displeasures are not as bad as free ones; on the second, unfree displeasures are worse than free ones. My view, again, is that this theory has little going for it. If lack of freedom affects the value of an intrinsic displeasure, there is little if any reason to deny that lack of freedom should also affect the value of an intrinsic pleasure.

7.6 On the Value of Worlds and Moral Obligation

What impact do these theories have on moral obligation, again, presupposing that, as of a time, we morally ought to bring about states of affairs that occur in the

[7] An even more complicated axiology would take into account, in addition to the influence of freedom and desert, the influence of, for instance, truth- or object-worthiness on the value of basics.

intrinsically best worlds then accessible to us? Assume that incompatibilism is true, and restrict attention to worlds in which no episodes of intrinsic attitudinal pleasure or displeasure (if there are some) are free. The gist of the robust freedom-sensitive theory is that unfree episodes of attitudinal pleasures and displeasures are intrinsically valueless. If this robust theory is true, the intrinsic value of worlds that are bereft of freedom is undermined. For if incompatibilism is true, no world atoms are true at these worlds. It follows that determinism undermines obligation if we ought to do the best we can and the robust view is the axiological basis for this sort of normative theory.

With the asymmetric theory, unfree attitudinal pleasures are less intrinsically good than free attitudinal pleasures, but freedom does not have any bearing on displeasures; an unfree episode of displeasure is just as intrinsically bad as it would be if it were free. Assuming that the pertinent world contains some episodes of attitudinal pleasure and some of displeasure, other things being equal, the effect of this freedom-sensitive asymmetric theory will be to diminish the overall intrinsic value of worlds. If we compare two worlds, *w1* and *w2*, that contain exactly the same raw amounts of attitudinal pleasure and displeasure but differ in that all the episodes of attitudinal pleasure in *w1* are free, whereas all of the pleasures in *w2* are unfree, the asymmetric theory implies that *w1* is intrinsically better than *w2*. Determinism will thus not threaten moral obligation if we ought to do the best we can and the asymmetric theory is true. But notice that in doing the best we can, we will be seeing to the occurrence of less net overall good than we would be if the episodes of attitudinal pleasure in the world were free. This is morally significant.

With the freedom-sensitive theories that are symmetric, the values of both episodes of intrinsic attitudinal pleasures and displeasures are adjusted according to whether the episodes are free or unfree. On the first version, unfree attitudinal pleasures are less good than free attitudinal pleasures and unfree displeasures are less bad than free displeasures. This theory's implications on the value of worlds are not so clear cut. This is because on this view, free pleasures and displeasures pull in opposite directions. Other things being equal, free pleasures augment the value of worlds; but (other things being equal) free displeasures depress the value of worlds. Consider three sorts of case. In the first, the world contains much more intrinsic attitudinal pleasure than intrinsic displeasure. If the first symmetric version is true, such a world (that is devoid of freedom) will be less overall good than a corresponding world in which the type- or near type-identical episodes of pleasure and displeasure are free. Determinism will, again, not subvert obligation. But it will still be true that, in doing the best we can, we will be bringing about less good than we would be in the corresponding "free world." In the second, the world contains much more intrinsic displeasure than intrinsic pleasure. If the first symmetric version is true, such a freedom-lacking world will be overall better than a corresponding free world. Determinism will leave obligation intact in this world. In the third, the freedom-stripped world contains roughly equal quantities of freedom-adjusted intrinsic attitudinal pleasure and intrinsic displeasure. If the first symmetric version were true, perhaps such worlds would be just as good as

they would be if they contained type- or near type-identical episodes of attitudinal pleasures and displeasures that were free. Moral obligation, even if determinism ruled the day, would remain secure.

On the second symmetric version, free pleasures are better than otherwise similar unfree pleasures, and free displeasures are not as bad as otherwise similar unfree displeasures. Thus, free pleasures and free displeasures are comrades, both contributing to the enhancement of the value of worlds. In each of the three sorts of case that we considered in connection with the first symmetric version, the second symmetric version yields the result that the unfree world is worse than its corresponding free world. Assuming the truth of incompatibilism, determinism will leave obligation intact. But in each of the three cases, in doing the best we can, we would be bringing about less good in the unfree world than we would be in the corresponding free world.

Finally, pain-adjusted freedom sensitive intrinsic attitudinal hedonism adjusts the values of episodes of attitudinal pain according to whether or not they are free and rates as intrinsically worthless episodes of attitudinal pleasure. Since I do not think this theory is plausible, I simply trace the implications of the first version regarding the value of certain worlds. On this version, unfree displeasures are not as bad as free ones. If this version is true, worlds with some attitudinal pleasure and some displeasure will be overall less intrinsically good than corresponding free worlds as will worlds with some attitudinal pleasure but no displeasure. In addition, worlds with some displeasure but no attitudinal pleasure will be overall more good than corresponding free worlds. But all these sorts of world will be *overall bad* although some will be less bad than others. Determinism will not undermine obligation at these worlds if we ought to do the best we can and the first version of the only-pain-adjusted theory is the correct axiology. However, at least as far as these three sorts of world are concerned, in doing the best we can, we will always be seeing to the occurrence of, that is, we will always be actualizing worlds that are overall intrinsically bad, something that would not be true if we were doing the best we can in corresponding free worlds. Again, this is morally significant.

7.7 The Value of Worlds and the Repugnant Conclusion

Assume that we have in hand a suitably developed freedom-sensitive version of SDAIAH and let us suppose that this is freedom-sensitive SDAIAH. Some might think that Parfit's attack on totalism undermines this world-ranking axiology as well. I first argue, though, that there are doubts about whether the repugnant conclusion undercuts totalism. I then show that principle *Best*, in conjunction with freedom-sensitive SDAIAH, reveals that inauthenticity of actional springs can have a definite impact on moral obligation: it is possible that one can fail to have an obligation to perform some action owing to that action's causally stemming from inauthentic actional springs.

7.7.1 An Objection to Totalism: the Repugnant Conclusion

For purposes of appreciating the force of Parfit's objection to totalism, it will be helpful to begin by showing that the simple theory (SIAH) falls prey to the problem of the Repugnant Conclusion. Parfit summarizes the problem in the following way:

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living . . .. As my choice of name suggests, I find this conclusion very hard to accept (1984, p. 388).

The apparent problem may be fleshed out in this way: assume, as the simple theory recommends, that net intrinsic attitudinal pleasure makes lives worth living; the value of a life is the total amount of intrinsic attitudinal pleasure in that life minus the total amount of intrinsic attitudinal displeasure in that life. Assume, further, that the value of a world is the sum of the values of the lives lived in that world. It would then follow that a highly populous world (HPW) in which each lives a life only just worth living is better than a less populous world (LPW) in which each lives a life high in welfare value. To illustrate, imagine that there are a billion billion people at HPW and each life in HPW contains no intrinsic displeasure but just one unit of intrinsic attitudinal pleasure. So each life in HPW has a value for the person of +1. Imagine, further, that LPW has far fewer people than HPW and each person in LPW enjoys a much higher level of welfare, +60, say, than does each in HPW. But suppose, assuming that the value of a world is the sum of the values of the lives (of persons) in it, the value of HPW far exceeds that of LPW.[8] It then seems to follow that HPW in which each lives a life that is barely worth living is better than LPW in which each lives a very good life. This result, as Parfit says, is repugnant. Parfit's alleged problem is a problem that Parfit seems to believe affects all totalistic world-ranking axiologies.

7.7.2 A Response to Parfit's Concerns Regarding Totalism

Feldman, however, questions whether the repugnant conclusion impugns all versions of totalism. He advances a solution to this problem that exploits Brentano's distinction between the value of an episode of something that is a primary intrinsic good for a person, and its value for the world. The value of an episode of intrinsic attitudinal pleasure (displeasure) for a person is equal to the amount of pleasure (displeasure) in that episode. The value of such an episode for the world is equal to its subject's desert-adjusted intrinsic value. As we previously explained, to calculate the worldly value of such an episode enjoyed by some subject, we consider both the subject's receipt level—the raw amount of intrinsic pleasure that the subject receives—and the subject's desert level—the extent to which the subject deserves to receive the pleasure that he does. Regard principles such as AXP1—–AXP5 as

[8] For simplicity, initially assume that all the recipients of value at a world are people.

giving us the desert-adjusted values *for the world* (as Feldman sees it) of episodes of intrinsic pleasure and displeasure that subjects receive. These principles, as we have seen, are part and parcel of SDAIAH, which is a totalistic world-ranking axiology. The proposed solution to the problem is now at hand. Using "Z" to denote the HPW, Feldman explains

> Let us assume that the residents of Z are relevantly like us—each of them is a person and starts out deserving a quality of life roughly equivalent to that which we deserve. To make the calculations simple, let us assume that each deserves a quality of life that rates +100 .... When a person deserves a life worth +100, but receives a life worth just +1, then that person has suffered a grave injustice. His receipt level is painfully lower than his desert level. As a result ... the pitifully small share of goods enjoyed by this person may be slightly good for him ... but it does not increase the value of the world at all. In fact ... we can say that the world is *made worse* by the fact that he receives just this good, since the total that he receives departs so dramatically from the amount that he deserves .... [T]he worldly level of each life lived in Z is approximately [let's assume] –49. There are a billion billion people living these low quality lives in Z. [So the totalistic desert-sensitive axiology that appeals to principles such as AXP1—AXP5] yields the result that Z has a worldly value of –49 billion billion. In short, Z is a horrible world .... Brentano's distinction [the value of an episode of a primary intrinsic good for a person vs. the value of this episode for the world] enables me to locate the precise spot where the reasoning behind the Repugnant Conclusion goes wrong. The mistake is in assuming that the value of the world as a whole is the sum of the values *for the residents* of the lives lived there. Following Brentano, my view is that the relevant assessment of the world is an assessment in terms of value *for the world*. When evaluated in the proposed way, taking due account of desert, it turns out that Z is indeed far worse than [the less populous world, LPW] (1997, pp. 209–210).

Some people, however, may not be convinced by this reply. Feldman himself considers this sort of objection:

> The critic might ask us to imagine a world, Z', in which there are a billion billion people. He might go on to ask us to imagine that each of them deserves +1 and gets +1. Given the axiological principles stated earlier [principles such as AXP-1—AXP-5], the worldly value of this world is two billion billion. Therefore the proposed Brentanist axiology implies that Z' is better than A. But it is not so. So the new axiology is erroneous (1997, p. 212, note omitted).

One of Feldman's primary responses to this brand of concern is the following:

> [I]t is not entirely clear that Z' ought to be considered horrible. Note that the residents of Z' are not like us. They deserve far less than we deserve. Each of them deserves just +1, and each of them gets exactly what he or she deserves. Since Z' is so incredibly populous, and since the total amount of good enjoyed by the residents is so huge, and since everything in Z' is said to be just as it ought to be, it is not clear that we should find Z' repugnant (1997, p. 212).

Some, though, may still find something amiss with this rejoinder. They might suggest that there could be two kinds of reason why the recipients of value in the HPW at issue (call it "HPW-2") deserve so little, whereas those in the LPW (what we may dub "LPW-2") deserve so much more. One is that they have "low capacities"; they are animals, pigs say, in which case, maybe, they could not be recipients of much value. These creatures could not possess or exemplify many of the kinds of value that humans can; so each of these recipients is getting all that it is capable

of deserving. Human beings, just in virtue of being persons, though, deserve much more. But it seems absurd to claim that a world containing, say, a trillion contented pigs and no humans is more valuable than a world containing a billion humans, all of whom are getting what they deserve. It is true that there is no injustice in HPW-2, but this world has lost the value of having persons. So the conclusion that HPW-2 is better than LPW-2 *is* repugnant.

I confess not to finding it obvious that this conclusion is repugnant. Suppose the pigs deserve to receive some pleasure and receive exactly what they deserve. Pig-world is perfectly just. I can imagine someone adding the following. "There are no human beings in pig-world; so much the better. The human race isn't exactly a stellar race. In fact, acquaintance with our history suggests that the human race is deplorable! The disappearance from a world of human beings may well be a blessing. Besides, why should the value of worlds supervene (even at least partly) on the kind of species that it has?"

One might continue in this vein: "suppose there is some way to assign value to different species in such a way that the value of a world is, at least in part, a function of the value of the species that populate that world. Suppose that the species, *Vulcan*, has very, very high "species value," far, far higher than that of the species *Homosapiens*. Let the HPW-3 world contain contented human beings and the LPW-3 world contain relatively fewer contented Vulcans. If you place so much stock in the human race, would you want to accept the conclusion (presumably delivered by the "species-value adjusted axiology") that LPW-3 is better than HPW-3? There is a further puzzle with the species-sensitive axiology. Just what determines the species-value of a species? If the relevant features are not "selected" correctly, it may just turn out that the species value of pigs is higher than that of human beings!"

People not satisfied with Feldman's rejoinder might insist that there is a second reason why it could be that the recipients of value in the pertinent HPW deserve so little, whereas those in the pertinent LPW deserve so much more. The people in the former world ("HPW-4") are capable of great things, and hence could deserve rich rewards. But they are lazy, feckless, and often wicked so that they only deserve a balance of one unit of pleasure over displeasure. To prefer a world with, say, a trillion fairly odious people, all of whom are getting the little they deserve to a world ("LPW-4") with, say, a billion virtuous high-achievers seems even more repugnant than the original repugnant conclusion. It would be to prefer a very populous purgatory to a less populous heaven.

Is LPW-4 indeed better than HPW-4? Although HPW-4 is filled with odious people and LPW-4 is filled with virtuous people (the "high achievers"), the worlds are *equally* just: each person in each world is getting his or her just deserts. The people in HPW-4 are not as nice as those in LPW-4. But, again, why should this have a bearing on the value of the *worlds*? If the worlds are equally just and one contains more pleasure than the other, why is it not *credible* to suppose that HPW-4 is better? If I were asked, "in which world would you prefer to live?", my response would be "in HPW-4." I would like to be around better people whether I myself were virtuous or morally depraved! (If I were a crafty morally depraved individual, I could exploit the virtuous.) But I see no reason why this sort of personal preference

has any bearing on the value of the *worlds*. The preference certainly does not imply that the virtuous world is better.

In sum, though the responses to Feldman's proposed solution that I have considered are instructive, I do not find them conclusive.

It is worth stressing that if one is convinced that LPW-4 (the virtuous world) is better than HPW-4 (the odious world), it is probably because one believes that virtue contributes to worldly value. Appreciating this fact, we can see that there is an easy recipe to generate a Parfittian-like objection against various totalistic world-ranking axiologies. Suppose one finds it plausible that some feature, *f*, such as *being virtuous*, or *being a species of a certain kind*, or *being knowledgeable* to which the world-ranking axiology in question is indifferent does in fact contribute to the intrinsic value of a world. Now simply select the two worlds, the HPW and the LPW, in such a fashion that the recipients of value in LPW but not those in HPW have feature *f*. If one is pretty much already convinced that having feature *f* contributes to worldly value, one will have generated Parfit's concern: in one's estimation, contrary to what the axiology under scrutiny implies, the LPW-*f* world is better than the HPW-(without *f*) world. This suggests to me, though, I may be wrong about this, that there is something pretty fishy about Parfit's objection. All one need do is pick on the feature that one antecedently believes contributes to worldly value and then "run" Parfit's objection in an attempt to confirm that the feature is indeed relevant to worldly value.

Why precisely, though, is this "fishy"? Well, Parfit's argument is designed to undermine *totalism*. Suppose that we have orchestrated Parfit's objection several times over, each time selecting a different feature that we believe contributes to worldly value (perhaps justice, freedom, species value, virtue, and so forth). Imagine that we have reached the "end of the line"; we do not think that there are *any* more features pertinent to worldly value. The resulting axiology will be highly complicated. It will adjust the values of its atoms for justice, freedom, and what have you. One wonders why the complex form of totalism that is the end result of such sequential modification would fall victim to Parfit's objection. To drive home his point, imagine that there is only *one* feature that contributes to worldly value and that feature is justice. (It would not make any substantial difference if we thought that there were several such features as long as the number is finite.) Then I cannot see why the claim that HPW-2 (pig-world) is better than LPW-2 or that HPW-4 (the virtuous world) is better than LPW-4 (the odious world) would be repugnant.

I suggest that the value of Parfit's interesting thought experiment involving the repugnant conclusion is that it helps both to call attention to, and to test, various candidates as prima facie plausible candidates for the sort of thing that affects worldly value. But I doubt whether the objection undermines totalism.

7.8 Inauthenticity and Obligation

Freedom-sensitive SDAIAH enables us to appreciate the bearing of freedom on the intrinsic value of worlds. We remind ourselves of the following:

(a) ***Free pleasure***: A pleasure is free only if it causally issues from authentic springs of action.
(b) An action-based pleasure (or displeasure) is a pleasure (or displeasure) that if deserved is deserved on the basis of performing some action.
(c) ***DAIP***: If an agent deserves an action-based pleasure (or displeasure), then that pleasure (or displeasure) causally arises—"it has its basis"—in springs of action that are authentic.
(d) A desert base fails to apply in a world if conditions at the world render it inappropriate to distribute goods on its basis. A desert base applies in a world if it is false that it fails to apply in this world.

7.8.1 Freedom's Bearing on the Value of Worlds

Imagine a world ("puppet-W") in which the sole desert bases that apply are action-implicating. Each recipient of value in this world is a subject (perhaps not human) who is globally manipulated by new-wave, super complex computers. When each value recipient takes pleasure or displeasure in some object, each is made to do so. The array of computers engineers into the subject's germane actional springs, thereby ensuring that no pleasures (or displeasures) that the subjects enjoy (or suffer) derive from authentic springs. It follows (from principles ***Free pleasure*** and ***DAIP***) that in this world, no pleasure or displeasure is free or deserved. The desert component of freedom-sensitive SDAIAH (together with the fact that no pleasure or displeasure in puppet-W is deserved) implies that each pleasure enjoyed and each displeasure suffered by each recipient of value in puppet-W contributes to worldly value in accordance with AXP3. This principle, recall, says that if one does not deserve any pleasure (or displeasure) but receives some pleasure (or displeasure), the intrinsic value of the episode of pleasure (or displeasure) that is received for the world is directly proportional to the amount of pleasure (or displeasure) it contains. The freedom component of freedom-sensitive SDAIAH implies that because no pleasures in puppet-W are free, their value is depressed in comparison with what their value would be if these pleasures were free; and this component implies, as well, that because no displeasures are free, these displeasures are worse than what they would be if they were free. We may conclude that puppet-W is (overall) good but not as good as the corresponding "free world," a world as "close" as possible to puppet-W but in which all the pleasures and displeasures are deserved or free.[9]

7.8.2 Inauthenticity and the Undoing of Moral Obligation

We may now dwell on the following issue. Posttransformation Beth's act of giving alms is derivatively inauthentic insofar as it causally arises from actional springs

[9] I assume that there are such worlds.

that are inauthentic. Can it be, in virtue of these antecedents being inauthentic, or in virtue of these inauthentic antecedents affecting something of direct concern to obligation, that Beth is not morally obligated to give alms? That she may not be obligated for this reason might seem pretty farfetched. If Beth is manipulated into destroying Ann's valuable manuscript, her destructive act issuing from engineered-in springs, it certainly seems that she has still done moral wrong though she may not be morally blameworthy for this wrongdoing. Similarly, although Beth may not be praiseworthy for giving alms, she may well be discharging an obligation in giving alms. Our intuitive judgments concerning such cases notwithstanding, I defend an affirmative response to our key query.

Recall, principle *Best* says that our obligations consist in making the world as good as we can:

Best: As of *t*, *s* ought to see to the occurrence of state of affairs, *p*, if and only if *p* is true at all the bests for *s* at *t* and *p* is not true at some worlds accessible to *s* at *t*.

More intuitively, according to *Best*, an act is obligatory for you if and only if you can do it, you can refrain from doing it, and (simplifying) it occurs in all the bests accessible to you. An act is permissible for you if and only if you can do it, you can refrain from doing it, and (simplifying) it occurs in some but not all the bests accessible to you. And an act is wrong for you if and only if you can do it, you can refrain from doing it, and (simplifying) it does not occur in any of the bests accessible to you.

In case "Tron-1," suppose that Tron is one of the manipulated inhabitants of puppet-W. The graduate director of an institution, on a certain occasion Tron has these options: at time, *t*, he can admit a conscientious student, Studious, or he can admit a lazybones, Slacker, or he can refrain from admitting either; admitting both is not an alternative. In a world otherwise as similar to Tron's but which is hospitable to desert, Studious would be much more deserving of admission than Slacker. We may, indeed, assume that in this as-similar world, at *t*, it is obligatory for Tron to admit Studious at some time t^*. In puppet-W, though, these considerations of desert are nullified. Given that matters of desert fall by the way, we may *coherently* assume that as of *t*, each world accessible to Tron—the one in which Studious is admitted, or the one in which Slacker is admitted, or the one in which neither student is admitted—are equal in worldly value; bear squarely in mind, for example, that although Studious, unlike Slacker, has exerted conscientious effort, in puppet-W Studious does not thereby *deserve* the action-based pleasures that she would otherwise deserve to receive if admitted. Generalizing, irrelevancy of desert in puppet-W results in an equalization of the values of all worlds accessible to agents, such as Tron, on various (but not all) occasions of choice.

Regarding the primary moral status of any of the three actions—whether it is morally right, wrong, or obligatory—that Tron performs at *t*, there are three salient options. On one option, each world accessible to Tron (at *t*) is best. It would follow on *Best* that none of Tron's alternatives is obligatory, none is wrong, and each is permissible. This "anything goes" outcome is implausible. On a second option, perhaps some might say that no world accessible to Tron is best. In this event, *Best* implies

that none of Tron's actions is obligatory, none is permissible, and each is wrong. Again, this "nothing goes" upshot is clearly untenable. A third option, preferable to the other two, is that each of Tron's actions is not morally right, wrong, or obligatory; rather it is amoral, lacking any primary deontic status. Thus, inauthenticity of actional springs can have a definite impact on obligation by way of having an impact on desert.

A less dramatic case ("Tron-2) that illustrates the possible influence of inauthenticity on obligation is one in which, given irrelevancy of desert, the worlds in which Tron admits a student are of equal value, and this value exceeds the value of the world in which Tron admits no student. In this case, it is permissible for Tron to admit either student. In contrast, in the as-similar world that accommodates desert, while it is obligatory for Tron to admit Studious, it is wrong for Tron to admit Slacker.

The bearing of inauthenticity on obligation is lost to us in worlds (such as the actual one) in which it is customary to assume that the actional springs of some recipients of value are authentic and those of others are inauthentic because these worlds are not "pure" worlds in which every actional spring is inauthentic. If Tron, for instance, is a subject of an "impure" or "mixed" world, then it is not the case that, on a relevant occasion of choice—for instance, on an occasion when Tron is considering which student, if any, to admit—each world accessible to Tron has a value equal in value to the value of any world then accessible to Tron. Impure worlds mask the pertinent effect of inauthenticity on obligation. Pure worlds such as Tron's expose this effect.

This sequence of reasoning that sustains the view that inauthenticity of actional springs can have a pronounced bearing on obligation involves a number of assumptions that may well be disputed. Two of these merit brief comment. The first is the assumption that we ought to what we do in all the best worlds accessible to us. The second is that this normative principle is to be paired with a version of a freedom- and desert-sensitive attitudinal hedonism as the world-ranking axiology. Many will reject these assumptions. But for those attracted to them, it is instructive to reveal that these assumptions commit one to acknowledging that there is a nontrivial connection between the authenticity of our springs of action, on the one hand, and moral obligation, on the other.

7.9 Conclusion

Summing up, many people have thought that determinism undermines moral responsibility. But it has been commonly assumed that determinism leaves the intrinsic value of worlds intact. A cardinal objective of this chapter has been to set this record straight. If we assume one of the simplest (and in my estimation) one of the most promising of world-ranking axiological theories, SDAIAH, then it is not so clear that lack of freedom does not have a bearing on the intrinsic values of worlds. This is because a credible version of attitudinal hedonism that tells us what makes one world better than another adjusts the values of either episodes of

intrinsic attitudinal pleasure or intrinsic displeasure, or both, among other things, according to whether or not the episodes are free or deserved. It follows that if what we morally ought to do at a time is what we do in all the intrinsically best worlds then accessible to us, lack of freedom may well have an indirect impact on moral obligation as a result of having a direct impact on the value of worlds.

Chapter 8
Hard Incompatibilism's Axiological Costs

8.1 The Issues

Pereboom, as we have chronicled, is an incompatibilist. Indeed, he defends a particularly demanding form of source incompatibilism that he calls "hard incompatibilism." To formulate this position we note, first, that an individual is an agent-cause if, roughly, the individual has the causal powers to produce actions, not by way of any agent-involving states or events, such as having beliefs and acquiring desires, but simply as a substance. An instance of agent causation would be an instance of some substance directly causing an event. According to hard incompatibilism, with the exception of agent-causal accounts of free action or moral responsibility, no compatibilist or libertarian account is true. A significant supplement to Pereboom's hard incompatibilism (also briefly documented previously) is that, contrary to what many have believed, a life devoid of free action would not be as detrimental as it has often been made out to be and, in certain respects, such a life would even be beneficial (Pereboom 1995, 2001, 2002). Pereboom, in effect, theorizes that fully appreciating hard incompatibilism reveals that this position is not such an exacting one after all.

In this chapter, drawing on some of the conclusions reached in the last chapter, I take issue with an element of the supplement. I argue that the hard incompatibilism is not as cost-free as Pereboom believes. Hard incompatibilism undermines aspects of desert; in so doing, it affects the intrinsic value of various worlds. Given its influence on worldly value, it should not be surprising that hard incompatibilism also bears significantly on moral obligation.

8.2 Hard Incompatibilism

Commencing with a brief overview of some of the key elements of hard incompatibilism, Pereboom argues that moral responsibility for an action depends primarily on the action's actual causal history and not on whether its agent could have avoided performing it. So in Pereboom's estimation, incompatibilism is not motivated by the thought that determinism expunges alternative possibilities without

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

the having of which no action can be free. Rather, Pereboom proposes that it is considerations concerning an action's causal origin or source that undergirds incompatibilism.

Elaborating, Pereboom submits that we can describe cases in which a victim of manipulation, such as a subject whose decisions are the product of antecedent elements—values, desires, and beliefs, for example—which neuroscientists surreptitiously implant in her, is not free and, hence, is not morally responsible for her behavior. (The Ann/Beth case is a case of this sort.) A causal history involving apt manipulation, a "manipulated causal history," undermines freedom and responsibility. In such cases, owing to the agent's not being the ultimate source of her action (mental or otherwise), the agent is not morally responsible for the action. More specifically, as I have explained, the pertinent principle to which Pereboom appeals to support his verdict of unfreedom and nonresponsibility in scenarios involving responsibility-subversive manipulation is as follows:

Principle O: If an agent is morally responsible for her deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control.

A deterministic causal history, Pereboom contends, is pertinently like a manipulated one: an action that is causally determined issues from sources—the distant past and the natural laws—over which the agent lacks any control. Contrary to compatibilists, Pereboom proposes that there is no relevant and principled difference between an action that results from responsibility-undermining manipulation and an action that has a more ordinary deterministic causal history. It follows that compatibilism is not sustainable (Pereboom 2001, pp. 89–126; 2002, p. 478).

Regarding libertarian accounts not wedded to agent-causation, Pereboom ventures that an indeterministic event–causal history, a history not including agent-causation, and in which various antecedents of an action, such as the agent's having of reasons, indeterministically cause elements in the action's etiology or the action itself, is not relevantly different from a manipulated one: this sort of history, also, undermines responsibility. This is because in scenarios involving indeterminism, just as in those involving determinism, antecedents over which the agent lacks any control produce the action. Again, Pereboom's position is that no relevant and principled difference can distinguish an action that results from responsibility-undermining manipulation from an action that has a more ordinary indeterminsitic causal history (Pereboom 2002, p. 478). He concludes that such libertarian theories are doomed to go the way of compatibilist ones.

Only agent-causation, Pereboom believes, allows for free action and moral responsibility. Agent-causation is coherent, but our empirical evidence does not support our being agent-causes. We, therefore, do not have the freedom that moral responsibility demands (Pereboom 2001, pp. 69–88).

8.3 First Cost: Hard Incompatibilism, Worldly Value, and the Repugnant Conclusion

Call any world devoid of agent-causation and in which either all events are causally determined or in which some events, including various mental actions such as (the having of) decisions, are indeterministically caused by proximate prior antecedents or concomitants, a "hard incompatibilist world." Appealing to the principle of ultimate origination (*Principle O*), we have just recorded that hard incompatibilists argue that all events in hard incompatibilist worlds, including our actions, are in the end the upshot of factors over which we have no freedom-level control: in determined worlds, our actions issue from the distant past and the natural laws; in agent-causal free indeterministic worlds, factors beyond our control contribute to the production but do not determine our actions, and there is nothing (such as agent-causation) that supplements the contribution of these factors to produce these actions (Pereboom 2001, p. 48). Hence, it follows from this fact and *Principle O* that the having of our desires or beliefs is not free. Abridging, we can say that according to hard incompatibilists, none of our desires or beliefs is free if our world is a hard incompatibilist one. Nor are any actions, mental or otherwise, that arise from such springs of action free in these worlds. This, in turn, supports the further hard incompatibilist conclusion that we are not morally responsible for any of our conduct in such worlds.

Should, though, hard incompatibilists concede that if our desires are not free in hard incompatibilist worlds, these desires in these worlds are *also* not authentic in the sense of "authentic" at issue? We have broached this sort of question before when we discussed source incompatibilist presuppositions (in Section 6.2). The answer given here mirrors the answer given previously to the relevantly analogous query. Hard incompatibilists are committed to the following assumption:

Hard Compatibilism—No Difference: There is no relevant and principled difference between an action whose causal history includes responsibility-undermining manipulation (as in the Ann/Beth scenario) and (i) an action that has a more ordinary deterministic causal history or (ii) an action that has a more ordinary indeterministic causal history.

If actions that derive from responsibility-undermining manipulation of the sort displayed in the Ann/Beth case, as hard incompatibilists should concede, are derivatively inauthentic, and if it is further assumed that more mundane deterministic or indeterministic causal histories are *not* relevantly different from a history involving this sort of manipulation, there is no basis to distinguish between the having of desires, the having of which is either deterministically or indeterministically caused, that is authentic, in the relevant sense of "authentic," and the having of such desires that is inauthentic. Thus, if hard incompatibilism is true, all desires in hard incompatibilist worlds are inauthentic.

Working with freedom-sensitive SDAIAH as the world-ranking axiology, we may appreciate a consequence of hard incompatibilism that many should,

presumably, regard as a cost of this position: hard incompatibilists run headlong into the problem of the repugnant conclusion.

Note, initially, what should be fairly obvious: hard incompatibilism can negatively affect worldly value by affecting freedom. Modify world puppet-W in this way: no agent in this world is a victim of manipulation but the world—call it "unfree-W"—is a hard incompatibilist world. Again, as in puppet-W, all the pleasures and displeasures in unfree-W are not free. So, just like puppet-W, unfree-W is (overall) good but not as good as the corresponding "free world," a world as "close" as possible to unfree-W but in which all the pleasures and displeasures are free.[1]

Now for hard incompatibilism's effect on desert, I have argued that an agent deserves an intrinsic attitudinal pleasure (or displeasure) on the basis of an action-implicating desert base only if that pleasure (or displeasure) causally arises—"it has its basis"—in springs of action that are authentic. I called such a pleasure an "action-based pleasure." According to hard incompatibilism, no springs of action in hard incompatibilist worlds are authentic. Hence, in such worlds it is false that action-based pleasures or displeasures are deserved. What we have just said about action-based pleasures and displeasures applies equally, mutatis mutandis, to other action-based primary intrinsic goods or evils (if there are any). It is in this fashion that hard incompatibilism undermines desert.

Confine attention to hard incompatibilist worlds (such as unfree-W) in which the only desert bases that apply are action-implicating. Maybe the recipients of desert in a world of this sort are not human, but this does not matter for our purposes. Even if an agent in such a world has exerted conscientious effort or has performed many morally good deeds or cultivated a stellar moral character and so is morally worthy, it is not true that because of these action-implicating desert bases, the agent deserves to receive (action-based) pleasures. Again, without authenticity, there is no foothold for desert of action-based attitudinal pleasures or displeasures. Assume that two worlds of the sort that Parfit envisions to launch the problem of the repugnant conclusion, the highly populous world ("HI-HPW") and the less populous world ("HI-LPW"), are both hard incompatibilist (possibly deterministic) worlds of this sort; these worlds are worlds in which only action-implicating desert bases apply. Then even freedom-sensitive SDAIAH ranks the former as intrinsically better than the latter. First, since no pleasures or displeasures in these worlds are free (given hard incompatibilism), freedom or its lack won't (relevantly) affect the values of episodes of intrinsic pleasures and intrinsic displeasures. Second, principle AXP3 implies that if one does not deserve any pleasure (or displeasure) but receives some pleasure (or displeasure), the intrinsic value of the episode of pleasure (or displeasure) that is received for the world is directly proportional to the amount of pleasure (or displeasure) it contains. Since no denizen of either world deserves (or, for that matter, deserves not) to receive any pleasure or displeasure (again, given hard incompatibilism), freedom-sensitive SDAIAH yields the same ranking of these two worlds as does the simple theory (SIAH): HI-HPW is better.

---

[1] Again, assume that there is such a world.

8.4 Second Cost: Hard Incompatibilism and Moral Obligation

Pereboom, we have seen, claims that although hard incompatibilism undermines the truth of judgments of moral responsibility, hard incompatibilism leaves unaffected the truth of judgments of moral obligation. But regarding hard incompatibilism's influence on moral obligation, matters are more complicated than might at first appear. Elucidating, assume, again, that principle *Best* is the correct account of (overall) moral obligation.

Best: As of *t*, *s* ought to see to the occurrence of state of affairs, *p*, if and only if *p* is true at all the bests for *s* at *t* and *p* is not true at some worlds accessible to *s* at *t*.

The two brands of hard incompatibilist world are agent-causal free deterministic worlds and agent-causal free indeterministic worlds. With respect to a spectrum of the former worlds, hard incompatibilists who accept freedom-sensitive SDAIAH will be saddled with the result that determinism undermines the truth of morally deontic judgments for a reason *other than* the reason that there is a requirement of alternative possibilities for moral obligation. To appreciate this reason, I assume that there is no such requirement and that the principle of doing the best we can is to be understood along these lines:

Best* (*Best* without alternatives): As of *t*, *s* ought to see to the occurrence of state of affairs, *p*, if and only if *p* is true at all the bests for *s* at *t*.

Imagine that a counterpart of Tron, Tron*, in unfree-W, is pondering (at time t^*) his options concerning which student, if any, to admit. (We remind ourselves that in this world, the sole desert bases that apply are action-implicating ones. Hard incompatibilists have it that in such a world no springs of action are authentic; so no action-based intrinsic attitudinal pleasures or intrinsic displeasures are free or deserved.) Again, assume that Tron* has three options: he can admit Studious*, or he can admit Slacker*, or can admit neither. We may suppose that, owing to the irrelevancy of desert in Tron*'s hard incompatibilist world, as of t^*, all the worlds accessible to Tron* have the same worldly value. Again, employing the line of reasoning that we previously did (see Section7.8.2), we may conclude that none of Tron*'s three alternatives is right, or wrong, or obligatory; each is amoral for Tron$*$. Thus, contrary to Pereboom's affirmation, it is not clear to me that obligation, right, and wrong remain unaffected in hard incompatibilist worlds that are deterministic.

Turning now to hard incompatibilist worlds that are indeterministic, we revert to the original version of the doing-the-best-we-can principle (*Best*) that entails that there is a requirement of alternative possibilities for obligation. Suppose, this time, that Tron* is an inhabitant of a hard incompatibilist *indeterministic* world in which the sole desert bases that apply are action-implicating ones. Again, hard incompatibilism implies that in such a world no springs of action are authentic and so no action-based intrinsic pleasures or displeasures are deserved. The line of reasoning employed previously (in Section7.8.2) yields the result that each world accessible to Tron* as of the relevant time—the world in which he admits Studious*, the world in which he admits Slacker*, and the world in which he admits neither student—has

the same worldly value. Now the argument unfolds in exactly the fashion in which we have seen before (Section7.8.2). Once again, there are three salient options. On one option, each world accessible to Tron* (at the relevant time, t^*) is best. It would follow on *Best* that none of Tron*'s alternatives is obligatory, none is wrong, and each is permissible. This is implausible. According to some, a second option is that no world accessible to Tron is best. Then *Best* implies that none of Tron*'s actions is obligatory, none is permissible, and each is wrong. Again, this is untenable. A third option, preferable to the prior two, is that each of Tron*'s actions is amoral. Thus, hard incompatibilism can have a pronounced influence on obligation by way of undermining desert.

8.5 Compatibilism, Worldly Value, and the Repugnant Conclusion

It is interesting to inquire whether a compatibilist who endorses freedom-sensitive SDAIAH runs into a similar problem concerning the repugnant conclusion as does the hard incompatibilist who endorses this world-ranking axiology. So let's imagine that some compatibilist account of free action and moral responsibility is true. Imagine, further, another highly populous deterministic world with a billion billion people and in which the only desert bases that apply are action-implicating ones. All of its recipients of value are children (though maybe not human children). Because of the horrible conditions, the children will die within a year of being born. Each child lives a life that is barely worth living, each receiving a life that is worth +1. Refer to this world as the "highly populous anguished world" (HPAW). The desires (and other springs of action) of the children in this world are not relationally authentic: as they die so young, none of these unfortunate children turns into an agent who will ever perform actions for which he or she is morally responsible. Thus, none of the desires of any of these children has the property of *being such that intentional actions that later issue from it are actions for which its agent is morally responsible*.[2] If an agent deserves an intrinsic attitudinal pleasure (or displeasure) on the basis of an action-implicating desert base, then that pleasure (or displeasure) issues from authentic springs of action. Since none of the springs of action of any of its residents is authentic, no (action-based) intrinsic pleasures or displeasures received by residents of this world are deserved.

Freedom-sensitive SDAIAH implies that the value of a world is the sum of the values, adjusted for both freedom and desert, enjoyed or suffered by its residents. The pertinent values of concern are the values of the intrinsic (action-based) pleasures and intrinsic (action-based) displeasures that the children in HPAW receive. Given their circumstances, it is false that each child deserves, in virtue of an action-implicating desert base, a much larger share of net intrinsic attitudinal pleasure than

[2] Its being true that the actional springs of these children are not relationally authentic does not (on the view I defend) imply that they are inauthentic. Rather, they are neither authentic nor inauthentic.

the share that each receives. It follows that even on freedom-sensitive SDAIAH, HPAW is a better world than a far less populous world such as HI-LPW. Let the less populated anguished world (LPAW) be a deterministic world that has fewer inhabitants than HPAW, all its recipients of value are children, each child will die by the time he or she is a year old, each receives a share of net action-based intrinsic attitudinal pleasure much larger than the share that each child in HPAW receives, but whose total worldly value is lower than the total worldly value of HPAW. Again, freedom-sensitive SDAIAH ranks this world as worse than HPAW. Thus, compatibilists, just like incompatibilists, confront the repugnant conclusion.

8.6 Conclusion

Hard incompatibilism may well be true. If it is, then it is not obvious that being without the freedom that moral responsibility requires is not without significant costs. I have suggested that whether living without free will is relatively cost-free, as Pereboom submits, is something that cannot be assessed independently of settling various highly complicated moral issues. One of them concerns desert. A second concerns the ranking of possible worlds in terms of their overall intrinsic value. Regarding the former, I have argued that hard incompatibilism compromises desert of action-based pleasures and displeasures. Regarding the latter, assuming a freedom- and desert-sensitive form of hedonism as one that determines worldly value, hard incompatibilism has a nonnegligible effect on the value of worlds. If we combine elements of both these things, hard incompatibilists incur costs having to do with the erosion of moral obligation in both deterministic and indeterministic worlds of certain sorts.

Chapter 9
Hard Incompatibilism, Practical Reason, and the Good

9.1 Introduction

In this chapter, I expose another axiologically related cost of hard incompatibilism. I assume that the hard incompatibilist accepts the view that determinism expunges alternative possibilities; if determinism is true, then no one can do otherwise than what he or she does. I argue, first, that there is a requirement of alternative possibilities for the truth of judgments of practical reason: persons can have objective, *pro tanto* reasons to do something only if they could have done otherwise. Thus determinism (and so hard incompatibilism) undermines the truth of judgments of practical reason. I propose, next, that, arguably, determinism precludes the obtaining of such states of affairs (or what we may refer to as "values") as *Simon's taking intrinsic pleasure in drinking wine is intrinsically good for Simon*, or *Al's taking intrinsic displeasure in sitting in the sun is intrinsically bad for Al*, and it does so owing to its undermining practical reason. If determinism does indeed preclude values (given certain incompatibilist assumptions), then there is a general argument, which I outline, for the conclusion that determinism undermines the intrinsic value of worlds; in so doing, determinism undermines moral obligation.

9.2 Practical Reason and Alternative Possibilities

Assume that the incompatibilist is correct in affirming that if free action indeed requires our being able to do otherwise—we are free in the germane sense of being free only if we could have done otherwise—"can" and its cognates, such as "could," are to be understood in the "categorical" (as opposed to the "conditional") sense: you could have done otherwise if and only if in a world with the same laws and the past as the actual world, at the pertinent time, you do other than what you do in the actual world.[1] In this section, I develop a line of argument for the view that lack of freedom to do otherwise subverts the truth of judgments of practical reason; thus, assuming

[1] On the conditional sense of "could," an agent could have done something other than what she did if and only if had she wanted, or tried, or chosen to do otherwise, she would have done otherwise.

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

that determinism precludes our having alternatives, determinism undercuts the truth of such judgments. We may commence by distinguishing judgments of practical reason from those of practical rationality. Practical reasons, roughly, are reasons to have our desires and goals, and to do what might secure these goals. So, for instance, when faced with certain options, we might ask which of them (if any) do we have most reason to do, or, which (if any) does reason forbid that we do, or which (if any) do we have a sufficient reason to do. Judgments of practical rationality (and irrationality), in contrast, relate to what one perceives or believes one's reasons for action to be. Thus, imagine that you (nonculpably) believe that you have most reason to take some pills because you (nonculpably) believe, on good authority, that the pills will alleviate your pain. However, you have decisive reasons *not* to take these pills because ingesting them would prove fatal. It would be practically rational for you to take the pills although you have no practical reason—"agent-external," objective, *pro tanto* reasons—to take them (the qualifier "practical" will, henceforth, be suppressed but assumed).

Rationality judgments are akin to responsibility judgments in one crucial respect: they are *agent-focused*. Both sorts of judgments are so focused in that, in the first instance, they concern an appraisal of the *agent*, and only derivatively, if at all, an appraisal of what the agent does (or fails to do), or an appraisal of the reasons that the agent has for behaving as she does. With responsibility judgments, when, for example, we hold an agent morally to blame, we are primarily *faulting the agent* and not what she has done. Such judgments disclose the moral worth of an agent with respect to some episode in her life—a person "expresses what she morally stands for" when she is morally responsible for some deed. When blameworthy, one expresses ill will (typically) toward another, and it is the expression of such ill will that (in part) sanctions the judgment that one's moral worth with respect to the relevant action has been diminished. On the view that I defend, an expression of such ill will is manifested when one (nonculpably) believes that some act is wrong, and despite this belief, performs the act; in short, when blameworthy, one acts in light of the belief that one is doing moral wrong. Similarly, when one is praiseworthy, one expresses good will (usually) toward another, and it is the expression of such good will that (in part) validates the judgment that one's moral worth vis `a vis the relevant deed has been augmented. When praiseworthy, one acts on the basis of the belief that one is doing what is obligatory or permissible (Haji 1998, 2002).

Similarly, with rationality judgments, when an agent acts irrationally, we are first and foremost faulting the agent; the agent is open to criticism or condemnation (which, of course, need not be moral) for doing (or not doing) what she does (or fails to do). When a judgment of irrationality is apt, we say to the agent things like "you were foolish, or crazy, or idiotic, or stupid for doing what you did." Even in cases involving akratic action, the primary target of appropriate judgments of irrationality is the agent. For instance, we fault the agent for not mustering sufficient self-control. Analogously, when an agent acts rationally, we are, in the first place, commending the agent, and again the commendation need not be moral.

Unlike such agent-focused judgments, judgments of practical reason are what we may call "act-focused" or, better perhaps, "nonagent-focused." When, for instance,

an act or an intentional omission does not have the support of reasons—perhaps there are no reasons for one to engage in the bit of behavior in question—what we are fundamentally faulting, it seems, is the bit of behavior, and only derivatively, if at all, the agent. Similar things are true when, for example, we judge that an action is morally wrong; the principal object of "deontic" appraisal is the action.

It may be thought that rationality judgments are closely associated with having practical reasons for or against doing something. Suppose that, on a particular occasion, we have several different *genuine alternatives*: holding "fixed" the past and the laws, we have different options. And suppose that our (objective, *pro tanto*) reasons to act in some way are stronger—perhaps far more so—than our reasons to act in any other way. Then we have most reason to act in this way; the reasons we have to act in this way outweigh the reasons we have not to act in this way. We may say that acting in this way is reason-wise obligatory; we reason-wise ought to act in this way. Suppose that, on a different occasion, we have sufficient or enough reason to act in two or more ways, and no better reason to act in any other way. Then we may say that each of these acts is reason-wise permissible; it is reason-wise permissible or right to act in either of these ways. Finally, suppose we have most reason not to act in a certain way. Then we may say that acting in this way is reason-wise wrong or forbidden (In more familiar language, we would say that we have decisive reason not to act in this way.).

Assume that when one is rationally at fault for something, one acts irrationally. This may not be quite right. For maybe it is possible, as I think it is, that one is less than fully rational, and so rationally at fault to some degree, with respect to what one does without being *irrational*. In the interest of simplicity, though, I ignore this possibility. Similarly, assume that if one is rationally commendable for doing something, one acts rationally. To draw out the alleged connection between rationality judgments and reasons in favor of or against doing something, one may now propose the following principles:

Irrationality Requires Reasons-Wise Wrongness (R1): If one is rationally at fault for doing something, *A*—if, that is, one is irrational with respect to doing *A*—then it is reason-wise wrong for one to do *A*.

The general thrust of *R1* is straightforward: despite one's having most reason not to do something, one still does it; so one is irrational—one is rationally at fault—for doing that thing.

Rationality Requires Reasons-Wise Obligatoriness (R2): If one is rationally commendable for doing something, *A*—if, that is, one is rational with respect to doing *A*—then it is reason-wise obligatory for one to do *A*.

Again, for simplicity, I am assuming that if one is rational with respect to doing something, then one has most reason to do that thing. (One could argue, for instance, for the alternative principle that one is rational regarding doing something only if one has most reason *or* sufficient reason for doing it.)

The germane parallelism between *R1* and *R2* and what we label *M1* and *M2*—roughly, responsibility correlates of *R1* and *R2*—is inescapable:

Moral Blameworthiness Requires Moral Wrongness (M1): If one is morally blameworthy for doing something, *A*, then it is morally wrong for one to do *A*.
Moral Praiseworthiness Requires Moral Obligatoriness (M2): If one is morally praiseworthy for doing something, *A*, then it is morally obligatory for one to do *A*.

I introduce *M1* and *M2* simply to draw out the following. One may think that just as moral blameworthiness presupposes moral wrongness, and moral praiseworthiness presupposes moral obligatoriness (or moral permissibility), so, irrationality presupposes reason-wise wrongness, and rationality presupposes reason-wise obligatoriness (or reason-wise permissibility).

I'm inclined to reject all four of these principles. However, if one *does* accept them, in particular, if one accepts *R1* and *R2*, then the hard incompatibilist will incur additional costs: given the hard incompatibilist assumptions introduced above, not only does determinism undermine the truth of judgments of practical reason, it undermines the truth of judgments of practical rationality as well.

Well, why might it be thought that determinism threatens the truth of judgments of practical reason? When it is moral obligation that is of concern, we may think of the "ought" implies "can" principle in this way: if you have most moral reason to do *A*, and, thus, if morality *requires* that you do *A*, then you can do *A*. (Again, understand "can" in the strong categorical sense of "can.") Suppose, now, that you have *most (practical) reason* to do *A*; as we said, *reason requires that you do A*, or that *you ought to do A from the point of view of reason*. Then it seems that you *can* do *A*. You cannot have an "obligation"—it cannot be *necessary*—from the point of view of reason to do something if you *cannot* do that thing. Imagine that you are in western Canada in Vancouver. Suppose you do not know how to swim, and you do not believe that there is a child drowning in the ocean off the Australian coast at the time you are enjoying your morning coffee. It turns out, though, that at this time a child *is* just about to drown off this coast. Had this child lived, she would have transformed the world from the predominantly horrible place it is into a blissful one. At the time you are sipping your coffee, how can you have a reason to save this child? For that matter, if the child were drowning at a time prior to the time at which you were enjoying your coffee, how could it be that, at the time you were having your coffee, you had a reason to save the child? And if it is false that you could have had a reason, at the time you were having the coffee, to save the child, it is also true that at this time, it is false that you could have had most reason to save the child.

One may object that in cases involving an action that one can perform only if one does something that it is physically (or nomologically) impossible for one to perform, or one does something that necessitates acquiring beliefs that one cannot (in one's circumstances) acquire, one *does* have a standing reason to perform the relevant action. It is just that this reason is overridden by others. For instance, even if it is physically impossible for you to save the child, you have a standing reason to do so, but this reason is overridden by another reason that you have: the fact that you cannot physically save the child.

Notice, first, though, that even when physical (or nomological) impossibility is involved, the following principle still seems to be true:

If one has a reason for doing something, A, that is not overridden by any others, then you can do A.

Second, the alleged overriding reason that you have for not saving the child is revealing. If you think that the (nonpractical) reason why your standing reason to save the child is overridden is that it is physically (or nomologically) impossible for you to save the child, why suppose that you have the standing reason at all? And this brings me to the third point: I don't see what, in the first place, motivates the view that one does have a standing reason to save the child (or, say, the Brazilian rain forest) even if it is physically (or nomologically) impossible for one to do so, or even if saving the child (or the rain forest) would necessitate acquiring beliefs that you cannot acquire.

So it seems that just as there is an association between the moral "ought" and "can," so there is a similar association between the "ought" of reason and "can." Indeed, the (moral) "ought" implies "can" principle appears to be a more restricted version of the general principle that

Reasons-Wise "Ought" Implies "Can" (KR): If you have most reason to do something, *A*, and, thus, if you reason-wise ought to do *A*, then you can do *A*.

KR's corollary is

Reasons-Wise "Ought Not" Implies "Can Refrain From" (KRC): If you reason-wise ought not to do something, *A*, then you can refrain from doing *A*.

Further, we should, I believe, accept this principle (*Reason-1*):

Reasons-Wise "Ought Not" amounts to Reason-Wise "Wrong" (Reason-1): You reason-wise ought not to do *A* if and only if it is reason-wise wrong for you to do *A*.

From *KRC* and *Reasons-1*, we derive:

Reason-Wise Wrongness Requires Alternatives (Reason-2): If it is reason-wise wrong for you to do *A*, then you can refrain from doing *A*.[2]

Reason-2 tells us that there is a requirement of alternative possibilities for *reason-wise* wrongness (just as there is such a requirement for *moral* wrongness).

We may now confirm that there is a requirement of alternative possibilities for *reason-wise* obligation (just as there is such a requirement for *moral* obligation). Here's the argument: if it is reason-wise obligatory for one to refrain from doing

[2] Again, expressions such as "it is reason-wise wrong for you to do *A*" may be thought to do too much violence to ordinary language—at least too much given that there are readily available alternatives such as "you have decisive reason not to do *A*." I ask the reader to put up with the strained terminology in the interests of appreciating the analogy I'm trying to highlight between the deontic predicates familiar in moral discourse and those predicates familiar in discourse about practical rationality and practical reason.

A, then it is reason-wise wrong for one to do *A* (from *Reason-1*). Further, if it is reason-wise wrong for one to do *A*, then one can do *A* (from the reason-wise "wrong" implies "can" analogue of (*KR*)).Therefore, if it is reason-wise obligatory for one to refrain from doing *A*, then one can do *A*.

If reason-wise wrongness and reason-wise obligation require alternative possibilities, then I see little reason to deny that reason-wise rightness, too, requires alternative possibilities.

If the truth of judgments of practical reason requires alternatives, but determinism effaces alternatives, then just as determinism threatens the truth of judgments of obligation, so determinism threatens the truth of judgments of practical reason. Further, suppose that *R1* and *R2 are* true:

(R1): If one is rationally at fault for doing something, *A*—if, that is, one is irrational with respect to doing *A*—then it is reason-wise wrong for one to do *A*.
(R2): If one is rationally commendable for doing something, *A*—if, that is, one is rational with respect to doing *A*—then it is reason-wise obligatory for one to do *A*.

Then, again, we may conclude that as determinism undermines the truth of judgments of practical reason, it undermines the truth of rationality judgments as well.

I have already recorded that I have reservations concerning *R1* and *R2* (as well as *M1* and *M2*). So even if there is a requirement of alternative possibilities for the truth of judgments of practical reason, as I have argued, determinism will not threaten judgments of rationality (or irrationality) on account of its being the case that one is rationally at fault for something only if it is reason-wise wrong for one to do that thing, and that one is rationally commendable for something only if it is reason-wise obligatory for one to do that thing. Nevertheless, that there is a requirement of alternative possibilities for the truth of judgments of practical reason has some sobering implications. I turn, next, to a brief discussion of one set of these implications that I find particularly troublesome.

9.3 Practical Reason and Value

I believe that there is a connection between something being good (or bad), whether intrinsically or extrinsically, for a person and practical reason, a connection that bears on whether determinism (again, given hard incompatibilist assumptions) precludes values. Assume that every episode of intrinsic attitudinal pleasure is intrinsically good, and every episode of intrinsic attitudinal displeasure is intrinsically bad. Assume, further, as we have previously done, that the objects of intrinsic attitudinal pleasure and intrinsic displeasure are states of affairs.

Suppose that *your taking intrinsic pleasure in drinking wine*—that is, *your being pleased in drinking wine*—is intrinsically good for you. If it is true that your being pleased in drinking wine is intrinsically good for you, then, it seems, you have a *pro tanto* reason—possibly a defeasible one—in favor of bringing about the state of affairs *your drinking wine pleases you*. Analogously, if your being displeased in sitting

in the sun is intrinsically bad for you, then you have a reason against bringing about the state of affairs *your taking intrinsic displeasure in sitting in the sun*. In a deterministic world, though, you have no practical *pro tanto* reasons—whether defeasible or not—to bring about any such states of affairs. Hence, in such a world, no states of affairs (that express values) relevantly like *your taking intrinsic pleasure in drinking wine is intrinsically good for you*, or *your taking intrinsic displeasure in sitting in the sun is intrinsically bad for you* obtain. I turn to two concerns with this thread of reasoning to the conclusion that determinism precludes relevant values shortly.

Very roughly, we may say that an actional state of affairs involving some agent is a state of affairs in which that agent engages in some sort of activity; she writes a paper, she drinks wine, she contemplates the sunset, she thinks about a mathematical problem, she forms an intention, she makes a choice, etc. Suppose, given some agent and some axiology (such as a version of hedonism), some actional state of affairs is either good, whether intrinsically or extrinsically, for that agent, or it is bad, whether intrinsically or extrinsically, for that agent.[3] Many people would agree that if it is true that *your drinking wine is good for you*, then you have a *pro tanto* reason in favor of the actional state of affairs involved in this state of affairs concerning goodness, to wit, *your drinking wine*. On this view, when something is in this sense good, this thing's goodness gives us reasons.[4] But, yet again, no one in a deterministic world has any such reasons. Thus, determinism precludes the truth of judgments such as *your drinking wine is good for you.*

Now, for the first concern regarding determinism's undercutting value, there are, of course, "buck-passers" about value (see, e.g., Scanlon 1998, p. 97; Parfit 2001, Stratton-Lake 2005). Suppose it is true that your drinking wine is extrinsically good for you. Buck-passers will probably agree that if your drinking wine is extrinsically good for you, then you have a reason in favor of *your drinking wine*. But they claim that it is not the fact that your drinking wine is (extrinsically) good for you that gives you this reason; goodness, as such, they insist, does not or could not give us reasons. Rather, they maintain that goodness is just the property of having other properties that provide reasons—in particular, reasons for pro-attitudes such as admiration. So, maybe, it is the pleasantness of the wine that provides the reason, say, to desire drinking the wine. It is not, on pain of circularity, the goodness of the wine that provides the reason since buck-passers see themselves as entirely doing away with goodness as something over and above reasons and the properties that provide them. Needless to say, the dispute between buck-passers and their rivals cannot be settled in this book. I do, however, believe that if a typical version of a buck-passing account of value is in the right ball park, then determinism *does* threaten values.

---

[3] I don't quite know what to say about intrinsic neutrality.

[4] It is possible to interpret this claim—that goodness gives us reasons—in a "benign" fashion. As Jonas Olson (2006, pp. 528–529) proposes, one might hold that you have a reason to x is to be analyzed, roughly, as your x-ing would be intrinsically good; this analysis of reasons in terms of intrinsic goodness is consistent with the view that reasons are provided by the natural properties that ground the goodness.

In broad strokes, buck-passers theorize that something has value (it is good, for example) if and only if it has "value-conferring properties" that give us reason to have certain attitudes toward it. Following Rabinowicz and Ronnow-Rasmussen (2004, p. 403; 2006, p. 114), I use "favoring" as a place-holder for various pro-attitudes and pro-responses. If the value-conferring properties provide reasons to favor the bearer of these properties for its own sake, the value of the bearer is intrinsic[5] ; if these properties provide reasons to favor the bearer for its effects or consequences, its value is instrumental. A "standardized" buck-passing account may be formulated as follows:

BPS: *x* is good if and only if *x* has value-conferring properties (other than its being good) that give us reason to favor *x* (either for its own sake or for its effects).

As various buck-passers make clear, the right–left implication is supposed to hold only for certain attitudes or responses. Rabinowicz and Ronnow-Rasmussen profess that they find especially attractive a pluralist version of a buck-passing account according to which "different kinds of value (admirability, desirability, etc.) correspond to different kinds of pro-attitudes that might be called for by the properties of the object" (2006, p. 114; see, also, 2004, pp. 400–401).[6]

Assume that some version of BPS is true. In particular, suppose

BPS1: *x* is intrinsically good if and only if *x* has value-conferring properties that give us reason to favor *x* for its own sake is true.[7] I've argued that if someone has a *pro tanto* reason to do something, then that person can (again, assuming the categorical understanding of "can") refrain from doing that thing. So if *x*'s value-conferring properties give you a reason to favor *x*—if you have such a reason—then you can refrain from favoring *x*. Assuming that determinism is true, and assuming that, on some occasion, you do favor *x*, you cannot, on that occasion, refrain from favoring *x*. It follows that you have no *pro tanto* reason to favor *x*. Generalizing, since we cannot have *pro tanto* reason to favor (or disfavor) anything if determinism is true and, hence, since nothing can give us *pro tanto* reason to favor it (or disfavor it) if determinism is true, nothing is intrinsically or instrumentally good if determinism and a buck-passing account of value such as BPS are true.

Some may want to exploit the buck-passing account's result that nothing is intrinsically (or instrumentally) good in a world in which we are unable to do otherwise than what we in fact do, to argue against this account. Imagine a world, *w1*, in which there is only one agent, Frank, who can adopt "value-relevant" attitudes toward objects that have what a buck-passing account would recognize as value-conferring properties. Assume that a certain object, O, has these properties, these

---

[5] There is a distinction between intrinsic value and final value. I agree, though, with Michael Zimmerman (2001, pp. 60–64) that all and only final value is intrinsic value.

[6] See, for instance, Vayrynen (2006) for concerns with this sort of value pluralism.

[7] Regarding BPS1, some buck-passers may opt for the view that something is intrinsically good only if it has properties that give us reason to favor it for its own sake.

properties give Frank a reason to favor O for its own sake, and Frank does indeed favor O for this reason. We are supposing, then, that *w1* accommodates reasons. Now imagine a world, *w2*, as close as possible to *w1* save that in *w2*, Frank cannot do otherwise for Frankfurt-sorts of reasons. In *w2*, Frank favors O because of what we are assuming are O's value-conferring properties but, unbeknownst to him, he could not have refrained from favoring O owing to some counterfactual intervener lurking in the wings (or some other failsafe mechanism). The intervener (or failsafe) mechanism does not in any way interfere in Frank's adopting the relevant attitude that he does toward O; the intervener (or mechanism) simply ensures that Frank can't fail to adopt this attitude. Then (as I have argued) Frank has no practical reason to adopt this attitude in *w2*. The buck-passing account generates disconcerting results in these two cases. If the account is true, then O is intrinsically good in *w1* but not in *w2*. O's value-conferring properties cannot give anyone reason to favor O in *w2* because this "Frankfurtized world" cannot accommodate reasons (again, given hard incompatibilist assumptions). But now one may suppose, and plausibly I think, that if O is intrinsically good in *w1*, it should be so in *w2* as well. In each of these worlds, we are supposing that O instantiates value-conferring properties; and in each of these worlds, Frank adopts the same sort of value-relevant attitude (he favors O). Thus, one might conclude that the buck-passing account is not true.

It may be objected that this thought experiment fails to cast doubt on the buck-passing account. If it calls into question anything, it calls into question the view that there is a requirement of alternative possibilities for practical reasons. More carefully, if Frankfurt-type examples impugn the principle of alternate possibilities—persons are morally responsible for what they have done only if they could have done otherwise—then such examples should also impugn:

Reason-Wise "Ought" Implies "Can Refrain From" (KKF): If you reason-wise ought to do *A*, then you can refrain from doing *A*.

My reply to this objection turns on the view that responsibility judgments and judgments of practical reason have different objects of evaluation; as I have explained, the first assesses persons, and the second actions or reasons. I said that appraisals of moral responsibility are, first and foremost, appraisals of the *person*—they are agent-focused. Regarding such appraisals, an agent can manifest ill will or good will by an action that she performs—she can disclose what she morally stands for on a particular occasion—even when she lacks, as she would in a Frankfurt-type situation, alternative possibilities. This is because such expression does not presuppose the availability of alternative options. Since an agent, in a Frankfurt-type case, discerns her situation and subsequently acts in just the way in which she would discern her situation and then act if there were no counterfactual intervener in the wings, in such cases agent-focused appraisals of responsibility should not be affected merely because the agent lacks alternatives. Or, coming at this point from another direction, if an act from your perspective is wrong—if you take yourself to be doing intentional wrong—then even if you lack alternatives (and, thus, your act is not in fact wrong), you can still express ill will in your conduct; and, similarly, if

you take yourself to be doing something that is obligatory, even without alternatives, you can still express good will in your conduct.

In contrast, appraisals of practical reason are pronouncedly *non*-agent-focused. They are *not* connected, in any necessary fashion, with how the agent perceives the situation or with the agent's germane motivations or beliefs. Whether an agent's act has the support or condemnation of practical reason—whether it has the property of *being reason-wise right*, or *being reason-wise wrong*, or *being reason-wise obligatory*—then, is *not* essentially associated with, for example, whether the agent performs that action in the belief that it has the support or censure of practical reason, or more generally, with the agent's perception of what she has reason to do or to avoid doing. Non-act-focused appraisals of practical reason (unlike those, for instance, of responsibility or rationality) turn primarily on whether the agent does indeed have reason to perform (or refrain from performing) the pertinent act.

If responsibility appraisals are different, in kind, from appraisals of practical reason, there is no reason to assume that the conditions requiring satisfaction for one to be apt must be the very conditions requiring satisfaction for the other to be apt. For instance, freedom to do otherwise may well be necessary for the latter sorts of appraisal but not for the former sorts. In light of this, there is little preliminary reason to believe that a counterexample against a proposed condition for the appropriateness of one sort of normative appraisal (for instance, Frankfurt-type examples as proposed counterexamples against the principle of alternative possibilities) will also be a counterexample against an analogous condition for the other. Hence, even though Frankfurt-type examples cast doubt on the principle of alternate possibilities, it is not clear that they cast doubt on the "reason-wise ought" implies "can refrain from" principle.

As for the second concern with the view that values are imperiled in a deterministic world (given incompatibilist assumptions) reflect, once again, on the following:

(1G)—If your drinking wine's pleasing you is intrinsically good for you, then you have a (possibly defeasible) reason in favor of bringing about *your drinking wine pleases you*.

(2G)—If your drinking wine is extrinsically good for you, then you have a (possibly defeasible) reason in favor of your drinking wine.

I have suggested that as determinism precludes our having practical reasons, determinism precludes values such as *your drinking wine's pleasing you is intrinsically good for you*, and *your drinking wine is extrinsically good for you*. One may object to this line of reasoning by offering these plausible "conditional" alternatives to (1G) and (2G):

Alternative (1G): If your drinking wine's pleasing you is intrinsically good for you, then you have a (possibly defeasible) reason in favor of bringing about *your drinking wine pleases you* if you can refrain from bringing about *your drinking wine pleases you*.

Alternative (2G): If your drinking wine is extrinsically good for you, then you have a (possibly defeasible) reason in favor of drinking wine if you can refrain from drinking wine.

These alternatives would allow for the existence of the relevant values even if alternative possibilities are absent, as they would be, in a deterministic world.

But now consider the following:

(1O): If you have an overall moral obligation to bring about some state of affairs, then you have a (possibly defeasible) reason to bring about this state of affairs.[8]
(2O): If you have an overall moral obligation to bring about some state of affairs, then you have a (possibly defeasible) reason to bring about this state of affairs if you can bring about this state of affairs

I find the unconditional thesis (1O) more plausible than its conditional counterpart. If unconditional (1O) is more plausible than conditional (2O), I don't think it too farfetched to suppose that the unconditional theses (1G) and (2G) are more plausible than their conditional counterparts (though, needless to say, this is contentious).

9.4 Practical Reason and Worldly Value

Imagine now that you are in a deterministic world. Assume that (1G) and (2G) are true.

(1G): If your drinking wine's pleasing you is intrinsically good for you, then you have a (possibly defeasible) reason in favor of bringing about *your drinking wine pleases you*.
(2G): If your drinking wine is extrinsically good for you, then you have a (possibly defeasible) reason in favor of your drinking wine.

And assume, too, that principle *Best* is true.

Best: As of *t*, *s* ought to see to the occurrence of state of affairs, *p*, if and only if *p* is true at all the bests for *s* at *t* and *p* is not true at some worlds accessible to *s* at *t*.

Assume, further, that the value of worlds is governed by a hedonistic world-ranking axiology such as SIAH, or SDAIAH, or freedom-sensitive SDAIAH. In this subsection, I argue for the view *that determinism undermines obligation as a result of undermining the intrinsic value of worlds*. (I retain the assumption that determinism precludes our being able to do otherwise, and the assumption that "can" is to be understood in the categorical sense.)

I begin by supposing that the simple theory, SIAH, is the correct welfare-ranking theory and that SIAH-W is the correct world-ranking theory. (In what follows in

[8] For simplicity, I've omitted temporal indices.

this section, I'll refer to either of these variations of the simple theory as the "simple theory.") On the simple theory, the value of a life or a world is the sum of the values of the atoms contained in that life or world. The atoms that SIAH recognizes, as we have seen, are relevantly similar to the following:

Positive Atom: At noon on Tuesday, October 16, 2001, Simon takes intrinsic attitudinal pleasure of intensity +8 in the fact that Simon drinks wine.
Negative Atom: At noon on Tuesday, October 17, 2001, Al takes intrinsic displeasure of intensity +3 in the fact that Al's cigarettes are damp.

We should, again as cautioned, be careful to distinguish between

Welfare Value (WLV): Simon's taking intrinsic pleasure in drinking wine is intrinsically good *for Simon*;

and

Worldly Value (WV): Simon's taking intrinsic pleasure in drinking wine is intrinsically good *for the world*.[9]

Assuming that the simple theory is both the world- and welfare-ranker (the world atoms that this theory recognizes are identical to the life atoms that this theory recognizes), if WV is true, then WLV is true as well. In a deterministic world, though, in which Simon exists, WLV is false for the reasons that I have explained. In such a world, WLV is true only if Simon has *pro tanto* reasons to bring about *Simon's taking intrinsic pleasure in his drinking wine*. He has no such reasons in such a world because he has no (categorical) alternatives in such a world. So WLV is false at such a world. It follows that WV is also false at such a world. More generally, no atoms that the simple theory recognizes as world atoms obtain at a deterministic world. The simple theory, then, construed as a world-ranker, declares that Simon's deterministic world has no intrinsic value.

Each world accessible to Simon—the one and only world—is devoid of intrinsic value. Such a world is neither best, nor is it worst, nor is its value any value between best or worst. I suggest (following the reasoning outlined in Section 7.8.2) that if *Best* is true, each of the actional states of affairs that Simon brings about in his deterministic world is not morally obligatory, not morally right, or not morally wrong for Simon; each is amoral for Simon in that each lacks a primary morally deontic status altogether.

What happens, though, when world atoms differ from life atoms? I confine attention to a single case: the case in which the desert-sensitive theory, SDAIAH, is the correct world-ranker and SIAH is the correct life-ranker. (It would make no difference to what is to come if freedom-sensitive SDAIAH were the correct world-ranker.) On SDAIAH, the world atoms are relevantly like the following:

World Atom: Simon takes intrinsic attitudinal pleasure (displeasure) of intensity *n1* and duration *m1* at time *t* in state of affairs *p* (e.g., *Simon's drinking wine*) when Simon deserves to degree *r1* to be taking that pleasure (or displeasure).

---

[9] There is an analogous pair concerning displeasure.

I introduce some terminology. Assume that a specific world atom is (simplifying)

W1: Simon's taking intrinsic attitudinal pleasure in drinking wine.

This *world atom's agent* is Simon, the person who takes intrinsic pleasure or displeasure in some state of affairs. Assuming, as we have, that the simple theory is the correct welfare-ranker—*W1's corresponding life atom* is *Simon's taking intrinsic attitudinal pleasure in drinking wine*. The *corresponding life atom's agent* is Simon as well.

Suppose a world atom, such as W1, is good for the world in which it obtains. Then its corresponding life atom is either good or bad for its agent. The former would be true, for instance, in a case involving neutral desert base or in a case in which the atom's agent, such as Simon, receives the pleasure that he deserves to receive. The latter might be true, for example, when the atom's agent receives the displeasure that he deserves to receive. Consider each of these possibilities—the corresponding life atom is good for its agent, and the corresponding life atom is bad for its agent—in turn. Regarding the first possibility, assuming that the world atom's agent occupies a deterministic world, we may argue as follows:

Possibility 1

(1GD) If a (specific, or better, perhaps, the specified) world atom is good for the world, then its corresponding life atom is good for its agent.
(2GD) If its corresponding life atom is good for its agent, then its agent has a relevant *pro tanto* reason.
So, for instance, if Simon's taking intrinsic pleasure in drinking wine is intrinsically good for Simon, then Simon has a *pro tanto* reason to bring about *Simon's taking intrinsic pleasure in drinking wine*.
(3GD) Therefore, if the specified world atom is good for the world, then its agent has a relevant *pro tanto* reason.
(4GD) It's false that (in a deterministic world) a world atom's agent has a relevant *pro tanto* reason.
(5GD) Therefore, it's false that (in a deterministic world) the specified world atom is good for the world.

Regarding the second possibility, instances in which though a world atom, such as W1, is good for the world, its corresponding life atom is bad for its agent, we may reason in an analogous fashion:

Possibility 2

(1BD) If the specified world atom is good for the world, then its corresponding life atom is bad for its agent.
(2BD) If its corresponding life atom is bad for its agent, then its agent has a relevant *pro tanto* reason.

Imagine a case in which the world is better in virtue of the fact that there is a perfect fit between the displeasure that some rogue receives and the displeasure that he deserves to receive, but in which the rogue is worse off because of the dis-

pleasure that he receives. If the rogue's taking displeasure in his being punished is intrinsically bad for the rogue, then the rogue has a *pro tanto* reason against bringing about his *being punished.*

(3BD) Therefore, if the specified world atom is good for the world, then its agent has a relevant *pro tanto* reason.
(4BD) It's false that (in a deterministic world) a world atom's agent has a relevant *pro tanto* reason.
(5BD) Therefore, it's false that (in a deterministic world) the specified world atom is good for the world.

Structurally similar arguments, with the necessary changes, can be formulated to show that no world atom in a deterministic world is bad for that world.

We may generalize. Suppose that the atoms of the world- and the life-ranker are linked in the following fashion: if a world atom is intrinsically good (or bad) for the world, then its corresponding life atom is intrinsically good (or bad) for its agent. Then we may avail ourselves of this argument:

(1) If a world atom is intrinsically good (or bad) for the world, then its corresponding agent has a relevant *pro tanto* reason. (This is, roughly, a possibly defeasible *pro tanto* reason in favor of or against bringing about the relevant state of affairs specified in the world atom's corresponding life atom.)
(2) If its corresponding agent has a relevant *pro tanto* reason, then it is false that determinism is true (as determinism precludes an agent's having *pro tanto* reasons).
(3) Therefore, if a world atom is intrinsically good (or bad) for the world, then it is false that determinism is true.

We have arrived at the following result. If SDAIAH is the world-ranker and SAIH the life-ranker, then no world atoms are true in any deterministic world including Simon's world. Simon's world is, thus, shorn of intrinsic value. Each of the actions that he performs in this world is, hence, amoral. (The string of reasoning to this conclusion assumes that there is a requirement of alternative possibilities for the truth of judgments of *pro tanto* reasons and that these alternatives are categorical.)

Even if one accepts the unconditional thesis (1O) about obligation,

(1O): If you have an overall moral obligation to bring about some state of affairs, then you have a (possibly defeasible) reason to bring about this state of affairs, but rejects the unconditional theses (1G and 2G) about values,
(1G): If your drinking wine's pleasing you is intrinsically good for you, then you have a (possibly defeasible) reason in favor of bringing about *your drinking wine pleases you*;
(2G): If your drinking wine is extrinsically good for you, then you have a (possibly defeasible) reason in favor of your drinking wine;

We still need to contend with the following alternative route to the conclusion that determinism precludes moral obligation (given incompatibilist assumptions): If you

ought to do something, then you have reason to do it. If you have reason to do it, then it is not the case that determinism is true. Thus, if you ought to do something, then it is not the case that determinism is true.

9.5 Practical Reason and Compatibilism

I conclude with some reflections on compatibilism, practical reason, and lack of freedom to do otherwise. Lying at the heart of prominent compatibilist theories concerning determinism and moral responsibility is a requirement (the "reasons requirement") that the choices—the decisions or intentions—or overt actions for which an agent is morally responsible be the outcomes of causal processes. Specifically, this requirement states that, for instance, a free decision be made *for reasons*, and its being made for reasons consists, partially, in its being caused nondeviantly by the agent's *having those reasons*. I have argued that there is a requirement of alternative possibilities for an agent's having *pro tanto* reasons. Thus, an agent who cannot do otherwise has no *pro tanto* reasons for any of her choices or actions. It follows that if, on the compatibilist theories in question, an agent's choice or action is free only if it causally derives from the agent's having of *pro tanto* reasons, then no choices or actions of this agent are free *if* she is unable to do otherwise. Some compatibilists may not be perturbed by this result because they embrace an account of "could have done otherwise" that is compatible with determinism. For other compatibilists, though, matters are more complicated. For these compatibilists, determinism's threat to moral responsibility, owing to determinism's impact on the truth of judgments of practical reason, cannot be so easily evaded.

John Fischer explains that semicompatibilism is the view that determinism is compatible with moral responsibility quite apart from whether determinism rules out the sort of freedom that involves access to alternative possibilities (Fischer and Ravizza 1998, pp. 52–53). Semicompatibilists wish to divorce their compatibilist theories from *any* requirement of alternative possibilities for morally responsible action. Fischer's elegant species of semicompatibilism, though, just like many other competing compatibilist views, is committed to the reasons requirement. Thus, on the construal of this requirement, which has it that a choice or action of an agent is free only if it causally issues from *pro tanto* reasons that the agent has, semicompatibilism is in jeopardy, given that there is a requirement of alternative possibilities for an agent's having *pro tanto* reasons.

In summary, I have argued that no judgments of practical reason (of the sort I have identified) can be true unless we have access to alternative possibilities. I have assumed that determinism is incompatible with our ever having done otherwise. I have further proposed that there is a connection between values (again, of the sort I have identified) and practical reason; without having *pro tanto* reasons to bring about certain states of affairs, there are no such values. Since determinism precludes our being able to do otherwise (on a categorical interpretation of "can"), determinism precludes our being able to bring about these states of affairs and, thus, determinism precludes such values.

Chapter 10
Value, Obligation, and Luck

10.1 The Issues: The Value of Worlds and Luck

Suppose, again, that our moral obligations consist in doing the best we can: as of some time, we ought morally to bring about the states of affairs that occur in the intrinsically best worlds then accessible to us. Suppose, in addition, that some version of attitudinal hedonism provides the axiology for this view of moral obligation. Suppose, finally, that what may be referred to as the libertarian freedom presupposition of obligation—the principle that one cannot, at a time, have a moral obligation to do something unless one has, at that time, a "genuine" alternative—is true. Then, to discharge the primary burden of this chapter, I argue that moral obligation is subject to luck because the intrinsic value of worlds is itself subject to luck.

The chapter is organized in this way: first, I clarify the libertarian freedom presupposition of obligation. Then I outline an archetypical account of libertarianism, and I address the so-called "luck objection" against such an account. Next, assuming the simple version of attitudinal hedonism (SIAH-W) as the world-ranking axiology, I argue that the intrinsic value of worlds is luck-infected because these values depend, in part, on what decisions we make. Given the libertarian freedom presupposition, these decisions that I take to be mental actions, in turn, are luck-infected. Obligation is, thus, also luck-infected if we ought to do the best we can. I then replace the simple species of attitudinal hedonism with the more complicated desert-sensitive variant (SDAIAH) that adjusts the values of episodes of pleasures and displeasures in accordance with whether subjects deserve these pleasures or displeasures. I argue that, on this axiology, obligation is once again luck-infected if we ought to do the best we can. I conclude with brief remarks on personal well-being and luck.

10.2 The Libertarian Freedom Presupposition of Obligation

Keep in mind that the normative theory that we are presupposing has, as its fundamental precept, principle *Best*: as of t, s ought to see to the occurrence of state of affairs, p, if and only if p is true at all the bests for s at t and p is not true at some worlds accessible to s at t. In Section 7.2, I exposed the bare bones of a complicated

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

line of reasoning for the view that moral obligation requires alternative possibilities. Briefly, I said that *Best* validates the principle that "ought" implies "can." But if this principle is true, "wrong" (and "right) imply "can" as well. *Best* also validates the principle that it is obligatory for one to refrain from doing something if and only if it is wrong for one to do that thing. This principle, in turn, in conjunction with the "wrong" implies "can" principle, yields the result that if it is obligatory for one to refrain from doing something, then one can do that thing. I concluded that the truth of judgments of obligation requires alternative possibilities: one has a moral obligation to do something only if one could have done otherwise. I interpret "could have done otherwise" in such judgments in a libertarian fashion: an agent could have done other than some action, *a*, that she did at time, *t*, only if, given the past up until she did *a*, and the laws, she could, at *t*, have refrained from *a*-ing. Abridging, we may say that obligation requires "genuine alternatives"; alternatives the agent has at a time even holding the past and the laws "fixed." The libertarian freedom presupposition of obligation can now be recast as follows:

Libertarian Freedom Presupposition of Obligation (LFPO): An agent, *s*, has a moral obligation to do *a* only if *a* is among *s*'s genuine alternatives (temporal indices have been omitted for simplicity).

LFPO entails that if, at *t*, *s* discharges *s*'s moral obligation to do *a* at t^* (t^* may be identical to or later than *t*), given the same past and the laws up until *s* does *a*, *s* could, at *t*, have done other than *a* at t^*.

10.3 Event-Causal Libertarianism

On a standard libertarian account of free action, an action is free only if there is some indeterminism at a point or points in the causal processes leading to that action. The standard account allows that an indirectly free action—an action whose freedom derives from the freedom of other actions to which this action is suitably related—may be determined by its immediate causal precursors. A directly free action is a free action that is not indirectly free. The account has it that even the immediate causal antecedents of a directly free action do not determine that action: given these antecedents, and the natural laws, there is some chance that the action not occur.[1] Such an account, then, is eminently suited to accommodate LFPO, the condition that moral obligation requires genuine alternatives.

In what follows, our concern is with directly free actions. Attention is confined to libertarian accounts that require that directly free actions be indeterministically caused by agent-involving *events*, such as the agent's having or recognition of reasons. We further restrict consideration to such "event-causal" accounts (what

[1] A recent defense of this sort of view is found in Kane (1996).

we referred to as "modest libertarianism" in Section 6.8) that eschew any appeal to agent- or substance-causation in their explication of directly free actions.[2]

It is uncontroversial that we receive some pleasures and displeasures because of the decisions that we make or, more generally, because of the actions that we perform. These are what we have called "action-based" pleasures (or displeasures). With this truism in mind, assuming the simple theory (SIAH-W) as the world-ranking axiology, and supposing some standard version of event-causal libertarianism as the account of free action, it appears that obligation is luck-infected because the intrinsic values of worlds are luck-infected. Their values, in turn, are luck-infected because these values depend, partly, on what decisions the agent makes, and these decisions, too, are luck-infected. I address, first, why and how decisions are luck-infected if a standard account of event-causal libertarianism is true.

10.4 Event-Causal Libertarianism and the Luck Objection

A simple illustration highlights a well-known problem of luck that, it has been argued, afflicts the sort of libertarianism at issue. Decisions, commonly conceived of as intentional actions, are frequently thought to be prime candidates for directly free actions. Suppose Fred is mulling over whether to tell the truth or to lie. He judges that, all things considered, it is better that he tell the truth, though reasons of self-interest tempt him to lie. He decides to tell the truth, and his having certain reasons to do so, including his making the all things considered judgment that he ought, on this occasion, to tell the truth, indeterministically causes this decision. Since this decision is so caused, there was a chance that Fred's deliberative process would have terminated in a decision to lie. Had he made this other decision, it would have been indeterministically caused by his having reasons of self-interest. *Everything* prior to the decision that Fred actually makes, including every feature of Fred, might have been just the same, and yet he could have made the alternative decision instead. To underscore this point, in the nearest possible world with the same past as the past in Fred's world, Fred's prior deliberations that mirror his deliberations in the actual world have resulted in the best judgment that it is better for him to tell the truth but Fred (or if we want, one of Fred's counterparts, Fred*) decides to lie instead.

Randolph Clarke, citing Mele, fills out the luck objection in this way: assume that were Fred to decide to lie, that decision would occur at the same time—time t—at which his actual decision occurs.

1. The actual world, where Fred decides at t to tell the truth, and some world W, where he decides at t to lie, have the same laws and do not diverge in their histories until time t.

---

[2] Randolph Clarke has argued that free action requires both event- and agent-causation. See, for example, Clarke (1996, 2003).

2. There is, then, no difference between these worlds to account for the difference between the decisions that Fred makes in the two worlds; nothing accounts for the difference between these decisions.
3. Hence, the difference at *t* between the actual world, where Fred decides to tell the truth, and world *W*, where he decides to lie, is just a matter of luck.
4. If the difference at *t* between the actual world, where Fred decides to tell the truth, and some possible world *W* with the same laws and the same predecision history, where Fred decides to lie, is just a matter of luck, then Fred does not freely make that decision in *W*.
5. Hence, in *W*, Fred does not freely decide to lie.
6. If, on these grounds, it is false that in *W*, Fred freely decides to lie, then it is also false that in the actual world, Fred freely decides to tell the truth.
7. Hence, Fred does not freely decide to tell the truth (Clarke 2004, p. 49).[3]

Unconvinced by this argument, Clarke first scrutinizes point (2), which he calls the "accounting claim." He says that the accounting claim is supposed to follow from the initial claim (1) regarding the indiscernibility of the relevant worlds prior to Fred's having made some decision and that it is meant to establish the subsequent claim (3) concerning luck. Clarke asks what interpretation of the accounting claim would underwrite its being true that this claim would both follow from the initial claim and help sustain the luck claim. He remarks that several advocates of the luck objection have proposed that on the relevant interpretation of the accounting claim, there is no explanation of why Fred decided to tell the truth rather than decided to lie. Hence, to "say that nothing accounts for the difference in question is to say that a certain contrastive explanation is unavailable" (2004, p. 50).

In Clarke's estimation, if this is how the accounting claim is to be understood, the claim is "a red herring" (2004. p. 50). It is a red herring, first, because Clarke argues that its truth is doubtful. Various accounts of contrastive explanation, such as Peter Lipton's, allow for such an explanation (Lipton 1990). Clarke writes:

> [S]uppose that because Fred judged that it would be better to tell the truth, it was much more likely that he would decide to tell the truth than that he would decide to lie [note omitted]. We have imagined that Fred's making this judgment was among the causes of his decision. Suppose it asked, by someone who took Fred to be strongly tempted to lie and who was unaware that he had made this judgment, why Fred decided to tell the truth rather than deciding to lie. Arguably, we could correctly answer the explanatory question in such a case, even given the imagined indeterminism. The explanation would cite Fred's judgment (or his judging) that it would be better to tell the truth. . . . The judgment rationally favored and made more probable than the alternative the decision to tell the truth, and it caused that decision. It was, then, a cause of the decision that bears an explanatorily relevant relation to that decision. Further, in the actual course of events, there was no occurrence that would

[3] In the original, there is a note after "truth" (in the conclusion 7). It says "Fred is my own invention; I have designed his case to instantiate the kind on which Mele initially focuses in addressing the luck problem. For the main claims that make up the luck argument, see Mele, *Free Will and Luck*, 7–8 and 59–60" (2004, p. 62). Mele develops the luck objection in a number of other works including 1999a,b; and 2005. Some of my reflections on luck appear in 2002, 2003, 2005a; and forthcoming.

> have borne that same relation to the decision to lie, had Fred decided to lie. Fred did not in fact judge that it would be better to lie, and no event made a decision to lie the more likely one (2004, p. 52).

Consider, though, Fred*s decision to lie. Given the facts of the case as stipulated, is there a plausible *non*contrastive explanation of this fact? Even Clarke appears to grant that it is important that there be such an explanation:

> If we ever act freely, then we can act for reasons when we act freely, and our free actions, when we so act, can be explained by citing the reasons for which we act. A highly credible view of such explanation of action has it that citing a reason explains an action only when the agent's having that reason caused the action. The rational explanation of action, on this view, requires that actions performed for reasons be caused by events—by the agent's having the reasons for which they act (Clarke 2005, p. 410).

Now it may seem that Fred*'s reasons of self-interest explain his making the decision to lie; these reasons indeterministically cause the decision. But as I have discussed in previous works, this is highly dubious or at least it seems not to be the full causal story (Haji 2000b, 2002, 2003, 2005a; forthcoming). The gist of the worry is not difficult to appreciate. We assume, as the event-causal libertarian insists, that Fred*'s decision is free. We assume, further, that it is a decision contrary to his decisive best judgment, the content of which is that it is better for him to tell the truth. It is not, then, an action—even if that—that manifests a breakdown of agency. Fred*s decision is, thus, a strict akratic action; it is a free, intentional action that is contrary to his consciously held best judgment. On customary accounts of akratic action, when an agent performs a strict akratic action, there is a misalignment between the motivational strength of the desire from which her action causally derives (the motivationally strongest desire) and her best judgment. Elaborating, Mele proposes that (i) best judgments normally are formed at least partly on the basis of our evaluation of the "objects" of our desires—what is desired and (ii) the motivational force of our desires does not always match our evaluation of their objects (Mele 1987, and 2004, p. 43). If both these things are true, then it should not be surprising that sometimes, although we decisively judge it better to *A* than to *B*, we are more strongly motivated to *B* than to *A*. Mele remarks:

> Thesis. . .[(i)] is a major plank in a standard conception of practical reasoning. In general, when we reason about what to do, we inquire about what it would be best, or better, or "good enough" to do, not about what we are most strongly motivated to do. When we ask such questions while having conflicting desires, our answers typically rest significantly on our assessments of the objects of our desires—which may be out of line with the motivational force of those desires, if thesis. . .[(ii)] is true. . . . Thesis. . .[(ii)]. . .is confirmed by common experience and thought experiments and has a foundation in empirical studies. Desire-strength is influenced not only by our evaluation of the objects of desires, but also by such factors as the perceived proximity of prospects for desire-satisfaction, the salience of desired objects in perception or in imagination, and the way we attend to desired objects. . . . Factors such as these need not have a matching effect on assessment of desired objects (2004, p. 243).

If, then, we accept typical accounts of akratic action, Fred*'s best judgment—that he ought to tell the truth—should stand opposed to his stronger desire—his desire

to lie. With Fred*'s libertarian free decision, though, we see no such misalignment because Fred* shares the relevant past with Fred. Stipulating that the past is fixed, Fred*'s desire to tell the truth does not differ in motivational strength from this desire of Fred's. But with Fred, we may assume that his desire to tell the truth has *greater* motivational clout than his self-interested desire to lie and, further, that there is no misalignment between this stronger desire and his judgment that it is better for him to tell the truth.

Perhaps a libertarian might claim that the misalignment in question does not occur *prior* to choice but at the moment of choice. The desire to lie, it may be suggested, does not become the strongest desire for Fred* until he makes it so at the moment of choice. On this view, it is false that akratic misalignment preexists the pertinent choice. Rather, such misalignment is created by the akratic agents themselves when they choose.[4] What, precisely, though, does the akrates do to shift the balance of motivational strength in favor of the desire that is allegedly out of kilter with his best judgment? Various explanations of the balance have been proposed, ones that make use of such things as selective focusing, failing to remind oneself how one will feel later knowing that one has acted contrary to one's better judgment, and conversely, thinking of how good one will feel afterward in light of being aware that one has successfully resisted temptation, and failing to make an effective attempt at self-control.[5] On the view that akratic misalignment occurs at the moment of choice, Fred*'s situation may be depicted in this way: we have said that *t* is the time at which Fred* makes the decision that he does. *Unlike* Fred, at *t*, Fred* selectively focuses, fails to make a concerted attempt at self-control, and so forth. At *t*, these activities (as we may say) of Fred*, in turn, bolster the motivational strength of his desire to lie, which, at *t*, causally gives rise to his decision to lie, despite his consciously held better judgment at *t* that he ought to tell the truth.

I limit comment to two concerns with this picture.[6] First, we have said that the actual world, where, at *t*, Fred decides to tell the truth, and possible world, *W*, in which he decides at *t* to lie, do not differ in *any* respects until *t*. It is, consequently, a mystery why, at *t*, Fred* engages in the activities that we have outlined when Fred, at this time, fails to engage in these activities. Engaging in these activities, or failing to do so, seems itself to be a matter of luck. Second, either these activities are free or they are not. If the latter, there is strong reason to doubt that the decision that Fred* makes is free because this decision stems from activities (that allegedly occur at the time when the decision is itself made) that are themselves not free.[7] If some of these activities, such as selective focusing, are not actions, if free, they will be indirectly

---

[4] See, for example, Kane (1999, p. 114, n. 7).

[5] For the first and third of these explanations, see Mele (1987, Chapters 5–6); for the second of these explanations, see, for example, Milo (1984, esp. Chapter 5).

[6] Other concerns are discussed in my 2005a, Section 2.

[7] I am not ruling out the possibility that if one event causes another, the two may occur simultaneously.

free. If they are actions—an intentional omission would qualify as an action—then, again, if free, they would presumably be indirectly free. If the latter—if, that is, these activities *are* free—and if it is false that free events are uncaused, as the event-causal libertarian assumes, these activities must themselves be indeterministically caused. Again, given that the actual world and world *W* are indiscernible right up until *t*, it is a mystery what the causal antecedents of these activities could be. More cautiously, if these activities have causal precursors, and Fred* engages in these activities, then again it seems to be entirely a matter of luck that Fred fails to engage in these activities.

In sum, the event-causal libertarian says that, consistent with the past and the laws being what they are, at *t* Fred can either freely decide to tell the truth, or at *t* he can decide to lie, and that whatever decision he makes at *t*, there is an action explanation of that decision. Roughly, apt reasons, it is claimed, indeterministically cause the decision that the agent makes. It seems that this view is over optimistic as Fred*'s scenario illustrates. We do not know the full causal story of Fred*'s akratic decision, and I strongly suspect that we do not know what it is because there is no plausible story to be told. It seems, then, that the accounting claim (2)—there is no difference between the actual world and possible world *W* to account for the difference between the decisions that Fred makes in the two worlds—is credible.

Clarke submits that the accounting claim (2) is a red herring for a second reason. He writes:

> What is at issue with the contrast argument–and what is at issue with objections from luck–is whether a certain agent who acted on a certain occasion acted freely. Whether an agent has acted freely in making a certain decision is a matter of whether she exercised the required control in making the decision. The control that is exercised in making a decision is an ontological matter. As causal theorists of action see it–and as any proponent of the type of event-causal libertarian account we are considering should see it–the control that is exercised is a matter of how and by what the decision is caused. If this view is correct, then once we know the full causal story, we know all that is pertinent to the question about control (though plainly we may still need to think about the information we have). Asking for an explanation is, of course, a way of getting at a causal story. But, it is important to note, in the case of the contrast argument, *we have the full causal story before we get to the accounting claim*. How then, can the accounting claim help us settle the issue of control? I doubt that it can (2004, pp. 53–54).

Suppose we grant, initially, that whether a decision is free is *solely* a matter of whether its agent exercised the required control in making it and that the freedom-level control that an agent exercises in making a decision is a function of how and by what the decision is caused. I have argued that the etiology of Fred*'s making the decision to lie is, at best, murky. Given the facts of the case, it is far from illuminating to rest with the contention that Fred*'s reasons to lie, perhaps in conjunction with his intention to make up his mind, causally gives rise to his akratic decision. Whether Fred* exercises the requisite control in making his decision for that decision to be free is, hence, questionable.

What of the assumption that the freedom of an action is solely a function of the control that its agent exercises in performing it? Assume that whatever decision Fred were to make, he would exercise freedom-level control in making it. Still, there is

no causal (or other) story to be told regarding why or how Fred makes his reasons to tell the truth "prevail," given the facts of the scenario, and similarly, no such story concerning why, under type- or near-type identical conditions, Fred* makes his reasons to lie "prevail."[8] Each exercises control, we are assuming, *in* making the decision that he makes but each has no control *over* which decision wins the day. In this sense, whatever decision is made, its being made is a matter of luck. I have suggested in other works (2003, 2005b; forthcoming) that perhaps an enticing implication of the luck objection is that, in addition to standard requirements of responsibility, and I now add of freedom, including the condition of freedom-level *control*, there is yet an additional "insulated-from-luck requirement." Some event-causal libertarians, at least, should not find farfetched the possibility of there being such a condition. This is because they accept the view that if an agent's decision causally issues from antecedents, such as desires and beliefs, that are not "truly the agent's own," then owing to these antecedents being engineered into the agent without the agent's knowledge or consent, even if the agent exercises *freedom-level control* in making the decision, the agent is *not* morally responsible for the decision (think back to the Ann/Beth scenario.) Of course, it is open to a libertarian to maintain that such a decision would be *free* but that its agent would not be *responsible* for it because some of its proximal causal precursors would not be authentic or because its agent would not be the ultimate originator of these precursors. It is, though, unclear why such a libertarian should shy away from the alternative view that freedom is complex, having a control component, an authenticity component, and perhaps, an insulated-from-luck component. Indeed, as we have registered, Pereboom (and others) champions the view that alternative possibilities are not required for free action but that the causal origins of an action are crucial to whether that action is free: to be free, an action must not causally issue from sources over which its agent has no control (Pereboom 2001).[9] This condition is distinct from the condition that if an action is free, its agent exercises freedom-level control—a variety of causal control—*in* performing it. If such a condition of ultimate origination finds favor among libertarians (and others), then perhaps the possibility of there being some insulated-from-luck requirement should not strike us as being outlandish.

To tie some ends together, according to the event–causal libertarian views in consideration, under normal occasions of choice, whichever directly free decision that an agent makes, the agent may well have freedom-level control in making that decision but lack control over why that decision rather than some other (or, perhaps, none at all) "prevails." Free decisions are, in this manner, luck-infected. One may, consequently, take the position that in virtue of being luck-infected, such decisions

---

[8] Mele attempts to provide such a story in Mele (2006, Chapter 5). For comments on this intriguing response to the luck objection, see my forthcoming, Chapter 9.

[9] Though not a libertarian, Pereboom shares this requirement with libertarians such as Bob Kane.

are not free. Or one may opt for the alternative that, despite being luck-infected, they are free.[10] I comment briefly on the coherence of this second position.

Clarke suggests that this (second) position is untenable. He explains

> [S]uppose. . .that in the actual world, Fred freely decides at *t* to tell the truth, and that in world *W*, Fred freely decides at *t* to lie. One further supposition: in all of the closest worlds (all those most similar to the actual world) where Fred does not freely decide at *t* to tell the truth, he freely decides at *t* to lie; were Fred not to freely decide at *t* to tell the truth, he would freely decide at *t* to lie. . . . Fred freely does something at *t* such that, were he to do it, it *would* be the case that at *t* he decides to tell the truth rather than deciding to lie. Fred is thus able to so act. And Fred is able to do something at *t* such that, were he to do it, he would do it freely and it would *not* be the case that at *t* he decides to tell the truth rather than deciding to lie. Then, the fact that at *t* he decides to tell the truth rather than deciding to lie depends on which of the things Fred is able to do at *t* he in fact freely does then. It is no stretch of the imagination to suppose that Fred is aware of this dependence. It would seem, then, that it is up to Fred whether, at *t*, he decides to tell the truth rather than deciding to lie. If that contrastive fact is up to Fred, then it is not *just* a matter of luck that at *t* Fred decides to tell the truth rather than deciding to lie. And if this contrastive fact is not just a matter of luck, neither is the difference between Fred's deciding at *t* to tell the truth and his deciding at *t* to lie, nor is the difference at *t* between the actual world, where Fred decides to tell the truth, and *W*, where he decides to lie. It thus seems that one cannot consistently accept the luck claim (3) and hold that Fred's actual decision is free and his alternative decision would have been, too (2004, p. 58).

A pivotal claim in this passage, however, that "it is up to Fred whether, at *t*, he decides to tell the truth rather than deciding to lie" is open to two (salient) interpretations. On the stronger, Fred exercises ("strong") freedom-level control (whatever this may consist in) in making the decision that he makes *and* he exercises such control, or control of this genre, *over* which decision "prevails." The weaker interpretation mirrors the stronger with the exclusion of the stronger's second conjunct. The conclusion that "one cannot consistently accept the luck claim (3) and hold that Fred's actual decision is free and his alternative decision would have been, too" seems to require the strong interpretation and not merely the weak one. I see little reason, though, to grant that Fred exercises strong freedom-level control in making the decision that he does. After all, at root, is the luck objection not an objection over whether the agent has such control in making the allegedly free decisions that he does?

It should be clear that of the two positions, first, whichever decision Fred makes, that decision is not free because it is luck-infected and, second, the alternative whose coherence Clarke questions, I favor the first. Which of these positions one ultimately sides with, happily, is not of direct concern in what follows on luck, worldly value, and obligation.

---

[10] See, for instance, Clarke (2005). Clarke's stance, more carefully, is this: if one takes it that whichever decision Fred makes, he makes that decision freely, then instead of the luck claim (3), one ought to hold that the difference at *t* between the actual world, where Fred decides to tell the truth, and world *W* where he decides to lie, is a matter of how Fred exercises his free will.

10.5 Obligation, Luck, and the Simple Theory

Reverting, now, to the issue of obligation's being luck-infected in virtue of the intrinsic values of worlds being so infected, suppose, first, that decisions may be free (on a standard event-causal libertarian account) in spite of being luck-infected. Assume (in the interests of simplicity) that Lucky is the sole recipient of value in Lucky's world. On a certain occasion, it is a matter of luck whether Lucky makes decision *d* or *d**. It is so in this way: given exactly the same past up until the time that he makes *d*, and the laws, *d** was in the offing. Lucky has no power in making it the case that *d* prevails rather than that *d** prevails. Imagine that if Lucky makes *d*, he receives an intrinsic attitudinal pleasure and no displeasure. If he does not make *d*, he makes *d** instead, and he does not receive any pleasure or any displeasure upon making *d**. If Lucky makes *d*, world *p* is actualized; if he makes *d**, world *w** is actualized. Stipulate that *p* and *w** have the same intrinsic value up until the time that Lucky makes one of *d* or *d**. Finally, assume that at the time, *t*, at which he makes one of *d* or *d**, *p* and *w* are the only worlds then accessible to him.

On the view that the simple theory (SIAH-W) is the world-ranking axiology, if Lucky makes *d*, the value of the world, *p*, that is actualized is greater than the value of the world, *w**, that would have been actualized had he made *d** instead. In this way, the intrinsic values of the accessible worlds depend (partially) on what decision Lucky makes. Their values are, hence, luck-infected. If Lucky makes *d*, then this mental action of his is obligatory: the state of affairs, *Lucky's making d*, occurs in all the bests accessible to him at *t* (and this state of affairs is not true at some world, *w**, then accessible to him); if he makes *d**, this action of his is wrong. Obligation is, thus, luck-infected as well.

I do not deny that obligation may be luck-infected for various reasons. In particular, some may stress that, assuming a standard version of event-causal libertarianism as the account of free action, Lucky's fulfilling his obligation at *t*, if he does fulfill his obligation at *t*, is luck-infected merely because the pertinent decision that Lucky makes at *t* is luck-infected. My emphasis is different: I underscore the view that obligation (on the relevant assumptions) is luck-infected because the *intrinsic values of worlds* themselves are luck-infected.

Some may have qualms over whether *d* can in fact be obligatory for Lucky when Lucky cannot control whether he makes *d* or *d**, although he has control *in* making whichever decision that he makes (see, e.g., Zimmerman 2005). I am uncertain whether this concern is cogent. Even if it is, the combination of event-causal libertarianism and the hedonistic axiology at issue preserves the result that the *intrinsic value of worlds* is luck-infected.

10.6 Unfreedom and Luck

So far, we have proceeded on the assumption that despite being luck-infected, the decision that Lucky makes at *t* is free. What, though, of the alternative? Assume that, because it is luck-infected, whatever decision that Lucky makes at *t*, that decision is not free. We may still, on this assumption, sustain the view that the value of

the *worlds* that obtain on Lucky's making whatever decision that he does is luck-infected. Again, if he makes *d*, and hence receives a pleasure, the world that is actualized is better than the world that would have been actualized had he made the alternative decision, d^*.

There is yet another way in which the value of worlds is luck-infected if it is assumed that the decision that Lucky makes is not freely made. We have argued that just as it is plausible to suppose that the value of an intrinsic attitudinal pleasure (for the world) is enhanced if (other things being equal) that pleasure is deserved, so (as I have proposed) it is credible to suppose that such value enhancement occurs if an intrinsic pleasure is *free*. A free intrinsic attitudinal pleasure contributes more to the intrinsic value of a world than an otherwise similar unfree pleasure. The story with intrinsic displeasures, as we have discussed, is more complicated. Two candidates we have mentioned are, first, that a free intrinsic displeasure is better than an otherwise similar unfree displeasure, and, second, that a free intrinsic displeasure is worse than an otherwise similar displeasure. Fortunately, the value of free or unfree displeasures is not essential to the point in development. If a pleasure is free, it is indirectly free, its freedom deriving from the freedom of those agent-involving events that are directly free. We have supposed that it is decisions that are directly free.

Assume, again, that if Lucky makes *d*, he receives a pleasure. If none of his decisions is free, as we are supposing, then this or any other pleasure (or displeasure) that he receives is not free. The world that is actualized upon his making *d* will have a certain value, a value not as high as it would have had had the pleasure that Lucky received when he made *d* been free. Now imagine worlds in which the decisions that Lucky makes can be free. Perhaps these are worlds where, in addition to Lucky's decisions being indeterministically event-caused, they are also agent-caused. Refer to these worlds as "free worlds." Worlds that do not accommodate free action may be labeled "unfree worlds." Luck is a matter of something's not being in one's control. Whether Lucky is born in a free or in an unfree world is something that is not in Lucky's control. It is, then, a matter of luck—"remote luck" we may say—whether Lucky is conceived in a free world or in an unfree world. Assume that a counterpart of Lucky, Lucky*, born in a free world, receives a pleasure upon making a decision, *dl*, that is free and that is type- or near type-identical to Lucky's decision, *d*. Assume that the unfree world in which Lucky makes *d* and the free world in which Lucky* makes *dl* are value-wise indiscernible up until the time that these decisions are made. Since it is a matter of remote luck in which of these worlds Lucky is born, the value of these worlds will, in this fashion, be luck-infected.

10.7 Obligation, Luck, and *Subject's Desert-Adjusted Intrinsic Attitudinal Hedonism*

It is straightforward to show that if SDAIAH (or freedom-sensitive SDAIAH) is the world-ranking axiology, and a standard version of event-causal libertarianism is the correct account of free action, then obligation is luck-infected because the

intrinsic values of worlds are, once again, luck-infected. Their values are luck-infected because these values depend, partly, on the luck-infected decisions that the agent makes. I simply draw the relevant implications on the assumption that indeterministically caused decisions, despite being luck-infected, are free and that SDAIAH is the correct world-ranking axiology.

Suppose, first, that if Lucky makes *d*, he receives an intrinsic attitudinal pleasure that is gratuitous and he receives no displeasure. It is not that if he makes *d*, he deserves to receive or not to receive any displeasure. If he does not make *d*, he receives no pleasure or displeasure. Further, it is not that he deserves to receive a pleasure or displeasure or he deserves not to receive these things if he does not make *d*. If he were to have received a pleasure or displeasure on not making *d*, that pleasure or displeasure would have been gratuitous. Even on the desert-sensitive axiology, SDAIAH, the value of the world that is actualized upon Lucky's making *d* is higher than the value of the world that would have been actualized had he made the alternative decision.

Suppose, second, that if Lucky makes *d*, he receives an action-based intrinsic attitudinal pleasure that is deserved. There is, moreover, a perfect "fit" between the pleasure that he receives upon making *d* and the pleasure that he deserves to receive because of making *d*. Suppose, also, that if Lucky makes *d*, he receives no displeasure, and it is not the case that he deserves or deserves not to receive a displeasure upon making *d*. If Lucky does not make *d*, he receives no pleasure or displeasure. It is also true that if he had not made *d* but had still received a pleasure or displeasure upon his not having made *d*, the pleasure or displeasure he would have received would have been gratuitous. Again, we may draw the relevant conclusion. SDAIAH generates the result that the value of the world that Lucky actualizes upon making *d* is higher than the value of the world that he would have actualized had he made the alternative decision instead. On either of these suppositions—whether the pleasure that Lucky receives is gratuitous or deserved, then, we may draw the conclusion that obligation is luck-infected.

10.8 Well-Being and Luck

I conclude with some remarks on luck and well-being. Assume that the simple theory provides the axiological account of when a person's life is intrinsically good for that person. Then Lucky's example forthrightly demonstrates one respect—an axiologically laden one—in which well-being is luck-infected. Again, suppose Lucky makes *d* as a result of prior agent-involving events indeterministically causing *d*. And, again, assume that upon making *d*, he receives an intrinsic attitudinal pleasure, something he would not have received had he failed to make *d*. All other things equal, Lucky's life is better in itself for him if, at the relevant time, he makes *d* rather than making some alternative. (We would derive similar results if some freedom-sensitive version of the simple theory were true.) Hence, how he fares is a matter of luck.

That personal well-being and obligation are luck-infected in the manner described is, to say the least, disquieting. The proposal that, given a standard incarnation of event-causal libertarianism, although the making of our pertinent decisions is luck-infected, these decisions may nevertheless be free is of little solace. For on these libertarian views, which decision "prevails" is not in your control—it is a matter of luck. Thus, how you fare and whether, on a certain occasion, you fulfill an obligation or do wrong is a matter of axiologically based luck. Not having the temperament of a gambler, I find little comfort in the thought that the luck may be good rather than bad.

Chapter 11
Freedom Presuppositions of Preferentism

11.1 An Alternative Axiology: Preferentism

The hedonist says that the *good life* is the pleasant life. The more net intrinsic attitudinal pleasure your life contains, the better your life is in itself for you. In past chapters, I have argued that freedom affects the values of intrinsic attitudinal pleasures and displeasures for the life of the recipient of these things. The attitudinal hedonist must, thus, take into account the impact of freedom on the values of life atoms; these are the positive and negative basic intrinsic value states that, on the axiology at issue, ultimately fix the intrinsic value of a life for the one who lives it. Attitudinal hedonism, as an account of well-being, has its rivals, the most prominent of which is perhaps preferentism. According to the preferentist (or desire satisfaction) account of well-being, your life goes well to the extent that you get what you want.[1] A primary aim of this chapter is to kindle support for the view that an association with freedom is an essential feature of any credible version of preferentism: roughly, any viable variety should include the requirement that it is the satisfaction and frustration of authentic desires—desires that are "truly one's own"—that contributes to well-being. If this is true, we have *some* reason to believe that the value of life atoms, no matter what the axiology in question, ought to be sensitive to freedom. A subsidiary aim is to show that if preferentism is the correct account of well-being, then source incompatibilists who are convinced that deprivation of a certain species of freedom undermines moral responsibility should also be convinced that deprivation of that variety of freedom threatens or at least has a nontrivial impact on well-being.

In this chapter, I first introduce a simple version of preferentism ("simple preferentism"). I focus on one objection (to the exclusion of sundry others) that has been raised against this version. I outline two possible responses to this objection, each of which motivates revisions to the simple version. Then I develop one of the revised versions ("subjective preferentism") and expose what even some

[1] Prominent recent defenders of preferentism have included Barry (1965, Chapters 10, 11); Rawls (1971, Chapter 7); Hare (1981, Chapter 5); Griffin (1986, Chapters 1, 2); and Raz (1986, Chapter 12).

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

preferentists may see as some of its shortfalls. This, again, prompts further amendments to the account. Revisions to subjective preferentism are made to yield "truth-sensitive preferentism." I then argue that even this sophisticated version of preferentism succumbs to a serious problem. It is this problem that exposes the connection between preferentism and freedom or authenticity. An outline of an account of yet another form of preferentism, "freedom-sensitive preferentism" is advanced. This account implies that whether a desire whose satisfaction or frustration contributes to well-being is free or not free affects well-being. Finally, I trace some of the implications of source incompatibilism for well-being on the assumption that some version of freedom-sensitive preferentism is true.

11.2 Simple Preferentism

The sorts of preferentist theories on which I focus say, roughly, that your life is going well for you to the extent that your desires—no matter what their objects—are satisfied rather than frustrated.[2] These are *satisfaction* rather than *object* versions of preferentism. According to object versions, something is good for a person if it is an object of his desires. According to satisfaction versions, something is good for a person if it consists in the satisfaction of one of his intrinsic desires. Actualist forms of satisfaction theories claim that the value of a life is determined by the extent to which actual desires are satisfied. Hypothetical forms imply that the value of a life is determined by the extent to which certain hypothetical (or would-be) desires are satisfied. I confine attention to actualist forms both because I think that they are stronger contenders than the hypothetical rivals and that if an acceptable version of preferentism must satisfy certain freedom requirements, any hypothetical version of the theory must satisfy such requirements as well.

To formulate simple preferentism, assume that desires are propositional attitudes. Whenever a person desires something, he desires it at a time, the object of his desire is a state of affairs, and his desire has an intensity (or strength). An intrinsic (or basic) desire that some state of affairs obtain is a desire that that state of affairs obtain for its own sake. On actualist and other versions of preferentism, it is the satisfaction and frustration of intrinsic desires that contribute toward how intrinsically well the person's life is going for that person. Assume, further, that a person's, *S*'s, desire that *p* is satisfied if *S* has this desire and *p* is true—the object of this desire obtains. It is no part of the theory that for some person's desire to be satisfied, the person must experience feelings of satisfaction. Adopting the following terminology will be helpful. An instance of *desire satisfaction* is a state of affairs in which a subject has an intrinsic desire (of a certain intensity) at some time that some state of affairs obtain and that state of affairs obtains. A *desire frustration* is a state of affairs in which a subject has an intrinsic desire (of a certain intensity) at some time

[2] Object versions are discussed by, among others, Wlodek Rabinowicz and Jan Osterberg (1996) and Ronnow-Rasmussen (2002a).

that some state of affairs obtain and it is false that that state of affairs obtains (see Heathwood 2005, p. 489; 2006, p. 540). On a simple, actualist desire satisfaction theory, the atoms of intrinsic value—the states of affairs that have their intrinsic value in the most basic, nonderivative way—are instances of desire satisfactions and frustrations. The intrinsic value of more complex entities, such as a person's life, is the sum of the value of the atoms contained in that life. The theory can be formulated in the following way:

Simple Preferentism

(i) Every instance of desire satisfaction is intrinsically good for its subject; every instance of desire frustration is intrinsically bad for its subject.

(ii) The intrinsic value for its subject of an instance of desire satisfaction is equal to the intensity of the desire satisfied; the intrinsic value for its subject of an instance of desire frustration is equal to –(the intensity of the desire frustrated).

(iii) The intrinsic value of a life (or a segment of a life) for the subject who lives it is equal to the sum of the intrinsic values of all the instances of desire satisfactions and frustrations contained in that life (or in that life segment).

Different objections have been directed against simple preferentism type theories, and various responses to them have been proposed. In the service of our aim to uncover the freedom presuppositions of preferentism, it is enough to dwell on what many take to be one of the most serious, the objection from remote desires. In broad strokes, the problem arises because simple preferentism implies that your life is made better for you when (other things being equal) an object of a desire that you have obtains regardless of what that object is (in some incarnations of the objection) or even when you have no idea that the object obtains (in other incarnations).[3] Sumner invites us to ponder an example in which Carl Sagan desires that we establish contact with extraterrestrial beings. Suppose a few thousand years from now, we do make contact with such beings. Sagan's desire is thus satisfied without his knowing that it is (Sumner 1996, p. 125). It would make little difference to the moral of the example if Sagan's desire were satisfied at a time when Sagan were still alive as long as Sagan were unaware that the intentional object of his desire obtained. Since the satisfaction of this desire has no discernible effect on Sagan, Sumner is dubious about whether its satisfaction could make Sagan's life better.

Krister Bykvist (like some others) adumbrates an interesting strategy to meet this response. We are to restrict preferentism in a manner in which it "should only count preferences that range over facts about the perferrer and her life" as welfare relevant; roughly, the objects of the pertinent desires must be states of affairs that essentially involve the preferrer (Bykvist 2002, pp. 480, 485, 488).[4] This strategy, as Bykvist is well aware, carries the burden of delimiting the relevant subset of desires.

---

[3] The objection has been discussed by, among others, Parfit (1984, p. 492), Griffin (1986, pp. 16–17), and Kagan (1998, p. 37).

[4] Overvold (1980), it seems, was the first to suggest this sort of move. Parfit (1984, p. 494) suggests a similar idea. Brad Hooker (1991) tries to improve on Overvold's contribution.

Sumner himself proposes an alternative (1996, pp. 112, 127, 128). He suggests that a state of affairs can make one better or worse off only if it enters or affects one's experience in some way (the "experiential requirement"). On one development of this suggestion, simple preferentism should be so amended that it implies that an agent's desire is satisfied only if the agent believes that the desire is satisfied. We turn to a subjective preferentism of this sort.

11.3 Subjective Preferentism

Chris Heathwood has recently advocated an interesting version of subjective preferentism (2006, pp. 547–551). Again, confining attention to elements of the theory pertinent to our discussion, Heathwood submits that an instance of *subjective desire satisfaction* is a state of affairs in which a subject has an intrinsic desire at some time for some state of affairs and believes at that time that that state of affairs obtains. An instance of *subjective desire frustration* occurs when the subject has an intrinsic desire for an object and believes that this object does not obtain (2006, p. 548). The value for the subject of an instance of subjective desire satisfaction is equal to the intensity of the desire that is the constituent of that instance. (Corresponding things are true for subjective desire frustrations.) If we replace each occurrence of "desire satisfaction" (and its cognates) with "subjective desire satisfaction" (and apt cognates), and we replace each occurrence of "desire frustration" (and its cognates) with "subjective desire frustration" (and appropriate cognates), in the formulation of simple preferentism, the result is subjective preferentism. On this view, your life is going well for you to the extent that you believe that the intrinsic desires that you have are satisfied rather than frustrated. Subjective preferentism escapes the objection from remote desires because it is only when a subject believes that the object of his intrinsic desire obtains that the pertinent state of affairs—his having an intrinsic desire that the object of his desire obtains—contributes to his welfare. When first contact is established, Sagan is not made better off because he is totally in the dark about the encounter with the aliens.

One feature in virtue of which preferentist approaches to well-being are attractive is that the subject receives what she wants. Our desires frequently express the aims that we deeply care about. We believe that fulfillment of these aims contributes to our well-being. Subjective preferentism, though, is compatible with the realization of none of the agent's primary aims; the agent can rack up well-being even though not one of her intrinsic desires is satisfied. Bykvist develops this concern in a poignant fashion. Though he is addressing Sumner's experiential requirement, his remarks apply to the subjective preferentist's demand (the "belief requirement") that a desire satisfaction contributes to well-being only if the subject believes that the object of the desire obtains. Bykvist writes

> The requirement entails that, even if you hate to be ridiculed, it cannot be bad for you to be ridiculed behind your back. Nor can it be bad for you to be deceived unknowingly, or to have false friends, people who only pretend to be your friends, even if you want to have true friends. These implications are vividly illustrated in the movie *The Truman Show*. This

> movie is about a man named Truman, who literally lives in a soap opera. From his birth he has been living inside a huge TV studio, surrounded by actors who have pretended to be his parents, friends, girlfriends, and neighbors. From the inside, his life has always looked perfect. He thinks he gets exactly what he wants. If we accept the experience requirement [or the belief requirement], we have to say that nothing bad happened to him (2002, p. 485, notes omitted).[5]

One may rejoin that Truman is not harmed by being deceived in this way because "what you don't know can't hurt you." However, some welfare theorists, including many inclined toward preferentism, may express concern that, just like certain forms of hedonism, subjective preferentism is a purely "mental state theory." On such theories, how well a person fares depends not even partly on, roughly, subject-external conditions that obtain in the world but solely upon her mental states. On the assumption that we do not want preferentism to collapse back into something like a pure mental state account, one can safely assume that such theorists will agree that there is something amiss with Truman's life (as Bykvist characterizes it).

Still, it might be proposed that if one does believe that there is something amiss with Truman's life, one may explain what one takes to be amiss by insisting that his life ranks low on *other* scales on which a life can be ranked, such as those that measure Truman's level of dignity (what happened to him surely degrades him) and perhaps his virtue (he leads a sheltered and controlled life, and so may lack certain virtues that require genuine risk and autonomy). So, in a sense, one can say that something bad has happened to Truman but just not in the welfare sense of "bad."[6] Although suggestive, this rejoinder is far from decisive. One can accept the fact that Truman's life ranks low on the sorts of scale mentioned consistently with adhering to the plausible verdict that Truman's welfare *is* compromised.

There is second problem with subjective preferentism. On this view, again, what contributes to an agent's welfare is that agent's *belief* in the satisfaction of an intrinsic desire of hers. If it is of no consequence to the promotion of one's welfare that the objects of one's intrinsic desires do not obtain, then why should it be of any consequence to enhancing welfare that one is not mistaken about the objects of one's intrinsic desires in the first place? You may think that you intrinsically desire that some state of affairs obtain when it is not the obtaining of this state of affairs but of some other that you really desire. Compare two cases. In the first, Rusty intrinsically desires that p obtain. Suppose that, though p fails to obtain, Rusty believes that it does. Subjective preferentism implies that (other things being equal), Rusty's well-being has been augmented. In the second, Dusty thinks that he intrinsically desires that p obtain when what he really desires is that q rather than p obtain. Suppose, just like Rusty, Dusty believes that p obtains. If Rusty racks up well-being (partly) in virtue of falsely believing that p obtains, why should it not be the case that Dusty also racks up well-being in virtue of his pertinent belief? "But Dusty doesn't have the desire that p!" one might protest. However, if accretion of well-being, as subjective preferentism

---

[5] For related literature, see Nozick (1974), Nagel (1979), and Griffin (1986).

[6] I owe this objection to Heathwood (personal correspondence).

implies, depends solely upon one's mental states and, further, false beliefs concerning the *obtaining* of objects of one's intrinsic desires are no bar to enhancing well-being, then having false beliefs concerning the *objects* of one's intrinsic desires should be no bar to enhancing well-being either. It seems, then, that if a preferentist thinks that you can rack up welfare simply by believing that the object of an intrinsic desire of yours obtains when it does not in fact obtain, there seems to be no principled basis for such a perferentist to deny that you can rack up welfare by believing that the object of an intrinsic desire that you falsely believe you have obtained. Refer to a preferentist of this second, more radical sort as a "doxastic preferentist." We can summarize the point just made in this way: if one is committed to "subjective preferentism, then, it seems, one lacks nonarbitrary grounds to avoid commitment to doxastic preferentism. But doxastic preferentism should strike even various preferentists as implausible. Suppose you falsely believe that you have an intrinsic desire that we make first contact. Suppose you believe, too, at a time when you still have this belief, that we have made first contact. Doxastic preferentism implies that your welfare has been enhanced in virtue of your believing that the object of a fictional desire obtains. Thus, doxastic preferentism would be an odd type of preferentism to hold because it falls victim to exactly the kind of problem involving things like deception or illusion that plague pure mental state theories, the very sorts of theory that, as many will see it, preferentism is designed to replace.

One may, of course, attempt to oppose the claim that commitment to subjective preferentism appears to bar one from nonarbitrarily resisting commitment to doxastic preferentism. It is not clear, though, how this sort of opposition would be successful without an explanation of why possessing the intrinsic desire itself, when the sort of preferentism at issue essentially associates well-being with pertinent *beliefs*, is necessary to well-being. In sum, if subjective preferentism is true, then provided one thinks that one has intrinsic desires for various objects and that these objects obtain, then (other things being equal) one's life should be highly good in itself for one. But even to many attracted to preferentism, this should be pretty hard to swallow. Hence, it may be recommended that subjective preferentism gives way to a more favorable version of the desire satisfaction theory.

11.4 Truth-Sensitive Preferentism

Assuming you are a preferentist who is convinced by these sorts of objection to subjective preferentism, a fairly obvious amendment suggests itself: in addition to the agent's believing that the objects of his intrinsic desires obtain (if this doxastic condition is indeed accepted), welfare accruement requires that these intentional objects in fact obtain. On the modified form of preferentism that is aptly sensitive to truth, the intrinsic value for a person of an instance of desire satisfaction is to be adjusted in accordance with whether or not the objects of this person's intrinsic desires obtain. Robust versions of this amendment imply that if one intrinsically desires something but that thing fails to obtain, then even if one believes that the

object obtains, the instance of subjective desire satisfaction is intrinsically worthless. Modest versions imply that the value for a person of an instance of desire satisfaction is augmented if the object of his intrinsic desire obtains and is mitigated if the object fails to obtain. On these moderate versions, truth enhances intrinsic goodness; lack of truth mitigates but need not eradicate intrinsic value. On either version, robust or modest, atoms of intrinsic value look relevantly like as follows:

Life Atom-2: Bob desiring at noon on Tuesday, October 6, 2005 with intensity +6 that Ann drinks a cold beer then; his believing then that Ann drinks a cold beer then; and Ann's drinking a cold beer then.

Generalizing, each atom has the following three parts: the desire that p be true, the belief that p is true, and p's actually being true. With moderate versions, to reflect the effect of truth on the value of such atoms, we could do the following. If the desire is satisfied, multiply the "raw intensity" of the instance of desire satisfaction (+6 for Atom-2) by 1; if the desire is not satisfied, multiply the raw intensity by 0.1. The result is the truth-adjusted intrinsic value of the atom. The value of a life for a person is equal to the sum of the truth-adjusted values of the atoms it contains. The theory can be formulated in the following way:

Truth-Sensitive Preferentism

(i) Every instance of desire satisfaction is intrinsically good for its subject; every instance of desire frustration is intrinsically bad for its subject.
(ii) The intrinsic value for its subject of an instance of desire satisfaction is equal to the truth-adjusted intrinsic value of that instance; the intrinsic value for its subject of an instance of desire frustration is equal to –(the truth-adjusted intrinsic value of that instance).
(iii) The intrinsic value of a life (or a segment of a life) for the subject who lives it is equal to the sum of the intrinsic values of all the instances of desire satisfactions and frustrations contained in that life (or in that life segment).

Truth-sensitive preferentism, robust or modest, however, suffers from a serious defect. Subjective preferentism and simple preferentism are deficient in this respect as well. To understand this defect, we begin with a general remark on well-being. Well-being concerns the good of the *agent*. If there is a transparent sense in which, though an agent has an intrinsic desire for something but the desire is "foreign" to, or not "really the agent's own," then even if the desire is satisfied, we would be hard-pressed to suppose that the *agent* is thereby made better off. In this way, *well-being is essentially personal*.

We have seen that global induction scenarios are revealing of when a person's desires may not be truly the person's own. In the Ann/Beth scenario in which Beth is engineered into being, in relevant respects, the psychological twin of Ann, we are inclined to the view that the implanted desires are not "truly Beth's own." We would not, for instance, agree that Beth is morally responsible for an action that causally derives from such desires.

Modifying the Ann/Beth case, suppose that this time around, aliens abduct Beth. Using their sophisticated medical technology, they prepare her so that she can be kept alive in a slimy womb; she is to live out the rest of her life in the Matrix. Once in her slimy bed of nutrients, Beth's brain is hooked up to a computer. At each instant, or during each appropriate temporal period, the aliens induce in Beth an intrinsic desire that she lives in the Matrix during that period, together with the (true) belief that she is then living in the Matrix. The abductors see to it that Beth possesses no other desires. A state of affairs, *p*, is all things considered good for a subject, *S*, if and only if the life *S* would lead were *p* to obtain is better than the life *S* would lead were *p* not to obtain. We may safely assume that satisfaction of Beth's induced and intrinsic desire to live in the Matrix is all things considered good. In this case (call it "Matrix-4"), as in the former Ann/Beth scenarios, Beth's engineered-in desires are not truly Beth's own. Hence, on the view that well-being is essentially personal, the satisfaction of these desires should not contribute to her well-being or (more plausibly, I believe) should not contribute as much to her well-being as they would have contributed had they been authentic.

Just as some may think that Matrix-like scenarios impugn forms of hedonism that are accounts of well-being, so others may think that such scenarios undermine preferentism.[7] This, however, is not my position. As I see them, the scenarios impel us toward the view that an acceptable account of preferentism should pay homage to some authenticity constraint.

This proposed moral, though, may not resonate with some readers. Why, it might be pressed, should satisfaction of intrinsic, engineered-in desires *not* enhance well-being (or at least not enhance it to the extent that satisfaction of otherwise similar authentic desires would)? It is important to bear in mind the presumption that there is a legitimate distinction between inauthentic desires and desires that are "truly the agent's own." Manifestly, Beth would not be morally responsible for her first few actions that expressed engineered-in springs. Appraisals of responsibility, as I have explained, are first and foremost the appraisals of the *agent*; they disclose the moral worth of an agent with respect, generally, to some episode in her life. When praiseworthy, a person's moral standing or "record" has been enhanced in relation to some germane episode; metaphorically speaking, a positive mark has been entered into her pertinent evaluative ledger. When blameworthy, an agent's standing or record has been blemished; a negative mark has been recorded into her ledger. Beth's actions, issuing from inauthentic desires, do not reflect either positively or negatively on *her.* How can they if it is true that Beth has had no say—roughly, she has had no control—in acquiring or sustaining the engineered-in springs, or in executing an intention or decision that issues from these springs, springs she would not have had but for the manipulation? Analogously, sometimes we deserve pleasure on the basis of our deeds. Even if manipulated Beth receives the action-based pleasure that she would have deserved to receive in virtue of performing otherwise similar deeds that had authentic desires, beliefs, and so forth as their causal antecedents, we

[7] Nozick's experience machine (1974, pp. 42–45) comes to mind.

would not think, as we previously emphasized, that manipulated Beth *deserves* the pleasure.

In the former case involving responsibility and in the latter case involving desert, most fundamentally, the view that Beth is not responsible or that she does not deserve the pleasure that she otherwise would have deserved rests on Beth's pertinent actions issuing from causal antecedents that are inauthentic. Why, then, should things be relevantly different when it is well-being that is at issue? Suppose you have an engineered-in intrinsic desire that you give alms to the poor and that this desire is satisfied. In one sense, the implanted desire is yours—you have it. But, being inauthentic, there is a sense in which it is not truly your own. If an intention that you form, or decision that you make, or action that you perform, on the basis of this sort of desire (and other engineered-in actional springs) is one for which you are not morally responsible, or if a pleasure that you receive on the basis of an action that nondeviantly issues from a desire of this sort fails to support your being deserving of this episode of pleasure, why should its satisfaction be thought to contribute to your well-being, or (more plausibly) to contribute as much to your well-being as its satisfaction would have contributed had this desire been truly your own? Perhaps the satisfaction of this desire contributes to your being happy. But it is one thing for your happiness to be enhanced; it is quite another for your welfare to be enhanced.[8]

The concern with truth-sensitive preferentism should now be evident: if truth-sensitive preferentism is true, Beth's life in Matrix-4 is good in itself for Beth. But this is not so or, minimally, it is highly controversial that this is so. Hence, truth-sensitive preferentism ought to be rejected in favor of a preferentist theory that is aptly sensitive to whether the desires whose satisfaction or frustration contribute to well-being are authentic or inauthentic.

11.5 Objections to an Authenticity Constraint and Replies

Prior to entertaining four objections (in this section) to this line of reasoning against truth-sensitive preferentism, I stave off a preliminary concern. When faced with preferentists who are Matrix bullet biters—they see hooked-up Beth in Matrix-4 as enjoying a high level of welfare—it might be worth asking them why they are *preferentists* rather than endorsers of a purely mental state theory, such as some brand of sensory hedonism, in the first place *if* they have such deep faith in the merits of subjective experience.

Now for the objections against an authenticity constraint, first, Chris Heathwood raises the following concern:

> If I acquire a desire by means of some artificial process (such as overzealous marketing, or brainwashing), it is tempting to think that satisfying the desire is not good for me

[8] Sumner (1996) claims that welfare consists in *authentic* happiness, the happiness of an informed and autonomous subject. See Section 11.5 below.

> because I don't "really want" the thing, because the desire conflicts with my "true self", because the desire is in some way "inauthentic". . . . [But] this will not undermine the actual desire-satisfaction theory. For the only sense that can be made of the idea of a desire being "inauthentic", or in conflict with one's "true self", is that the desire conflicts with many other desires held by the person. And if a desire conflicts with many other desires held by a person (in that satisfying it means frustrating others), the desire is (or could be) all-things-considered defective: the life the subject would lead were the desire not satisfied is worse than the life he would lead were it satisfied (2006, p. 494)**.**

However, Heathwood's chief claim—"the only sense that can be made of the idea of a desire being "inauthentic", or in conflict with one's "true self", is that the desire conflicts with many other desires held by the person"—is mistaken. There are a number of accounts of authenticity that speak to the claim's not being true or at least being highly controversial. A simple hierarchical account of authenticity says that an agent's (first-order) desire is authentic if and only if the agent identifies with it. A person may fail to identify with a first-order desire simply because he has no second-order volition concerning it and *not* because he has conflicting first-order desires. Or consider Mele's alternative. Ignoring various nuances of his view, Mele proposes that a desire is not authentic if it is acquired by a process that bypasses ordinary mechanisms of deliberative control (1995, pp. 166–168, 171–172, 183–184; 2006, pp. 166–167, 170–173.). A desire acquired in this fashion need *not* conflict with any other desire held by the person. (Yet another account of authenticity that sheds doubt on Heathwood's claim is the one previously outlined in Chapter 5.) Thus, Heathwood is not justified in claiming that the only way to understand inauthenticity of desire is in terms of conflicts among desires held by the subject.

A second objection first directs attention to the distinction between welfare-relevant factors, on the one hand, and world- or life-relevant factors, on the other. We explained before that welfare-relevant factors are factors that affect the intrinsic value of a life *for a person*. World- or life-relevant factors are factors that affect the intrinsic value of *a world* or *a life*. The objection avers that the negative verdict concerning the intrinsic value of Beth's life for Beth in Matrix-4 conflates these two sorts of factor. The *life* that Beth leads is a terrible one; no one, for instance, would *recommend* a life of this sort to anyone. Still, *Beth's life is good in itself for Beth* because her intrinsic desires are all satisfied. Factors such as truth and authenticity, it is further proposed, are world- or life-relevant, and not welfare-relevant factors.

Since I have addressed a similar sort of objection directed against, for instance, object-worthy attitudinal hedonism, my response will be somewhat truncated. The kernel of the concern is that it is not evident what makes some factor world-relevant but welfare-irrelevant. Why, for example, can it *not* be the case that truth and authenticity are both world- and welfare-relevant factors? We previously entertained the view that it is internal indiscernibility that undergirds the distinction between the two sorts of factor. Suppose Dusty desires that his peers respect him; he believes that this desire is satisfied but the sad truth is that his peers think little of him. Dusty*, one of Dusty's counterparts, has the same sort of desire that Dusty has and this desire is in fact satisfied. We may suppose that Dusty's and Dusty*'s pertinent life

segments in which they have these desires are internally indiscernible; they "feel" or "appear" exactly the same from the "inside." The desired moral one may want to draw from all of this is that truth is, thus, not welfare-relevant. Similar things could be recommended about authenticity.

Once again, invoking the notion of *affect** will permit us to make some headway (again, the asterisk simply flags a semitechnical sense of "affect"). Suppose we want to know whether truth is welfare relevant; we want to know whether a desire satisfaction in which the object of the desire obtains is welfare-relevant. We compare the segment of an agent's, *S*'s, life in which *S* has the pertinent intrinsic desire and in which the object of that desire obtains with an otherwise type-identical or near type-identical segment in which *S* has this desire but in which its object does not obtain. If the two segments are internally indiscernible, then the factor in question does not affect* *S*. It may be suggested that the following proposal—"Proposal WR"—is true:

Proposal WR: Factors that do not affect* *S* are not welfare-relevant.
Since truth does not affect* its agent, truth is not welfare-relevant and, similarly, with authenticity.

Proposal WR, though, is mistaken. We can show that it is erroneous in a slightly different way from the way in which we formerly showed that it is defective. Many of our desires and beliefs are unconscious. *S*'s desire *D* that *p* is weakly unconscious if and only if *S* does not know or believe that *S* has *D*; similarly, *S*'s belief *B* that *q* is weakly unconscious if and only if *S* does not know or believe that *S* has *B*. *S*'s desire *D* (belief *B*) is strongly unconscious if and only if *D* (*B*) is weakly unconscious, and apart from outside help or careful self-scrutiny, *S* cannot come to know or believe that *S* has *D* (*B*).[9] *S*'s desire is *active* if this desire need not first be aroused (or "activated") in some way to exert an influence on *S*'s conduct (Mele 2003, pp. 30–31). An *occurrent** desire of *S* is, among other things, active *and* present to *S*'s consciousness—*S* is aware that *S* has this desire. An occurrent desire of *S* is, among other things, active but need not be present to *S*'s consciousness. Some weakly unconscious desires, as well as some strongly unconscious ones, are active, just as occurrent* desires are; they require no prior arousal to exert influence on the agent's conduct. Other unconscious desires are, presumably, not of this sort; they require activation prior to being able to have any bearing on conduct.

Suppose Dusty has some deeply unconscious active desire *D* that *p*, *p* obtains, and Dusty has a deeply unconscious belief that *p* obtains. If we compare the segment of Dusty's life in which Dusty has these deeply unconscious elements with an otherwise type-identical (or near type-identical) segment in which Dusty lacks these elements, the two segments are internally indiscernible. Proposal WR yields the result that an intrinsic (active) desire's being deeply unconscious is welfare-irrelevant. But this is false, on any workable version of preferentism. Many of our unconscious active desires, just as many of our occurrent* desires, are, presumably, authentic,

[9] For pertinent discussion on unconscious belief, see Audi (1982).

and many of our unconscious active desires, just like many of our occurrent* ones, have a pronounced influence on such things as our emotional states, dispositions to act, and overt action. It thus appears that a credible version of preferentism cannot, in an *un*principled fashion, restrict the desires whose satisfaction contribute to welfare merely to occurrent* desires. It should thus not matter, on a preferentist theory, whether our intrinsic desires are strongly, weakly, or not unconscious provided that these are active. Hence, proposal WR is erroneous.

It may be thought that a workable version of preferentism should exclude from the set of intrinsic desires that contribute to well-being *any* dispositional desire. Heathwood has advanced this intriguing line of reasoning for such exclusion: if a preferentist theory includes merely dispositional desires and beliefs as welfare relevant, we encounter the problem of there being an infinitely many of these things. This is because, he thinks, coinciding in each of us at every moment are infinitely many desire/belief pairs toward the same proposition (here is one of them: *that I will not be killed by a falling asteroid in 3 seconds*). Without the restriction, the theory would entail that everyone's life is equally good or bad.[10]

Two comments are in order. First, as Mele argues, it seems not to be the case that every so-called dispositional (standing) desire is a bona fide desire. Mele directs attention to Robert Audi's proposal that not all dispositions to believe occurrently that *p* are beliefs that *p* (Audi, 1994). The following example illuminates Audi's proposal:

> Before reading this sentence, you had a disposition to believe occurrently that a girl who has 127 apples has more fruit than a boy who has 52 oranges. But that disposition is very different from your standing belief that Los Angeles is in California. That Los Angeles is in California is something that you learned, that you occurrently believed at times in the past, and that you remember. Your relation to the fruity fact is of a very different kind. Presumably, you encountered that fact here for the first time (Mele 2003, p. 31).

Mele then argues that if not all dispositions to believe occurrently that *p* are beliefs that *p*, then, it appears, not all dispositions to desire occurrently that *p* are desires that *p*:

> A year ago, Sam put a house he inherited on the market, wanting to get at least $75,000 for it. He still wants that, and he has rejected all lower offers. Sam is disposed to want occcurrently to sell the house for $75,099, is disposed to want occurrently to sell it for $75,108, and so on. If Sam were offered any of these amounts for his house, he would want occurrently to sell it for that amount. But it certainly seems to be stretching things to claim that Sam has correspondingly numerous desires—even standing ones—to sell the house. His relevant standing desire, if he has one, is to sell his house for at least $75,000, and he is disposed to desire occurrently to sell it for at least that much. However, the numerous related dispositions I mentioned are not plausibly counted as numerous desires. (2003, p. 31)

So perhaps your so-called dispositional desire that you will not be killed by a falling asteroid in 3 seconds, just like Sam's so-called dispositional desire to sell

[10] Personal correspondence.

his house for $75,108 is not a bona fide desire. One may, thus, have reservations concerning the claim that each of us has infinitely many dispositional desires.

Second, Mele helpfully suggests that it is sufficient for an agent's having a standing (i.e., dispositional) desire for *x* that the agent frequently has had occurrent desires for *x*, these occurrent desires manifested an ongoing disposition to have such desires, and the agent still has that disposition (2003, p. 33; 2007). Imagine that Sam's desire to sell his house for at least $75,000 satisfies these conditions and that this desire is inactive. (The housing market has crashed; Sam has had the desire for a long, long time, etc.) Even though inactive (and not present to Sam's consciousness), once activated this desire can exert significant influence on Sam's behavior. Hence, it is questionable whether viable versions of preferentism should exclude even such desires from the set of desires that contribute to well-being, and this means that truth-sensitive preferentism again has no principled way of restricting the set of welfare-related desires so that it can avoid problematic cases like Matrix-4.

A third objection to there being an authenticity constraint on preferentism is that the notion of *a desire's being authentic* is murky; it may reasonably be demanded that one advance an analysis of authenticity. I took steps toward doing so in Chapter 5. As I acknowledged, though, some might be skeptical about the relational account that I constructed. Even so, the following remark merits emphasis. Regarding moral responsibility, there is widespread agreement that, for instance, one cannot be blameworthy for an action unless one has control in performing that action; the action is free. But, nevertheless, there is considerable disagreement over the analysis of the germane concept of *being free*. Similarly, there is a *fair* measure of consensus that in cases such as global induction ones, the agent fails to act on desires that are truly her own or authentic though there is disagreement over the analysis of authenticity. Given this measure of agreement, and the reminder that well-being is essentially personal, it may be enough to draw attention to the fact that simple preferentism, subjective preferentism, and truth-sensitive preferentism, whatever their merits or other defects, are not sufficiently discerning of the authenticity (or freedom) presuppositions of well-being, whatever the correct analysis of authenticity.

Before entertaining a fourth objection, we should highlight central features of a freedom-sensitive preferentism. We can say that someone authentically desires something if and only if she desires that thing and this desire is authentic. On *freedom-sensitive preferentism*, the atoms of value all look pertinently like this:

Atom 2: Bob's authentically desiring at noon on Tuesday, October 6, 2005 with intensity +6 that Ann drinks a cold beer then; his believing then that Ann drinks a cold beer then; and Ann's drinking a cold beer then.

The intrinsic value of such atoms may be adjusted for truth in the manner already discussed. According to this view, if a person has many authentic, intrinsic desires, believes (at appropriate times) that these desires are satisfied, and these desires are (at apt times) in fact satisfied, then her life is going well for her. To the extent that her authentic, intrinsic desires are frustrated (other things being equal), her life is not going so well for her.

Maybe some preferentists, who look favorably upon moderate versions of truth-sensitive preferentism, will protest that just as lack of truth does not render desire satisfactions intrinsically worthless, so lack of authenticity should not be taken to render desire satisfactions intrinsically valueless. We should opt, instead, for the modest position that authenticity or its absence affects the intrinsic value of desire satisfactions: the satisfaction of authentic desires enhances the intrinsic value of a life for a person but to a lesser extent than would the satisfaction of an otherwise similar inauthentic desire. Similarly, the frustration of an authentic desire contributes more to the intrinsic badness of a life than would the frustration of an otherwise analogous inauthentic desire. I am not sure how such a modest variety of freedom-sensitive preferentism could be secured if well-being is essentially personal. Here, I merely call attention to this possible option.

In summary, I propose that each form of preferentism should incorporate an authenticity requirement to the effect that the values of the desire satisfactions and frustrations that contribute to well-being are to be adjusted to reflect whether the agent's pertinent desires are authentic or inauthentic.

Turning, now, to the fourth objection, this objection is directed against a freedom-sensitive version of preferentism that incorporates the relational account of authenticity that I have advanced. At the heart of this objection is the idea that the concept of *a desire's being normatively wise authentic* differs, contrary to what I have assumed, from that of *a desire's being* (what we may refer to as)*"welfare-wise authentic."* An example involving a pair of cases will help to flesh out the objection. In the first case, Ken, struggling to quit smoking, enlists the help of a team, members of which implant in Ken, among other things, progressively stronger desires to avoid smoking. Ken eventually succeeds in kicking the habit. In a second, Ken*, also wanting to quit, refuses help from the team. However, overly zealous members of the team surreptitiously work on Ken* implanting in him the cluster of apt psychological elements by the processes that bypass Ken*'s capacities of deliberative control. The implanted actional antecedents are, thus, inauthentic. Ken*, just like Ken, gives up smoking. If Ken's well-being is augmented as a result of the psychosurgery, Ken*'s well-being should be augmented as well, or so the objection goes, and it should be augmented despite its being the case that the germane antecedents of action that are implanted in Ken* are all inauthentic. Authenticity of desires, it is concluded, has nothing essential to do with well-being.

I advance three responses. First, in pertinent regards the Ken/Ken* case is somewhat analogous to ones advanced against forms of preferentism, such as subjective preferentism, that are not sensitive to truth. Suppose Abe believes that his intrinsic desire for something is satisfied when the object of this desire does not in fact obtain; and suppose that Abe*, who is as close to Abe as possible in all relevant respects, also believes, though this time truly believes, that his type- or near type-identical desire is satisfied. Focusing on the germane life segments of these individuals in which they have these desires, even though the segments are alike in all other respects, a preferentist who is inclined to the view that truth has a bearing on well-being will find nothing amiss with the judgment that Abe's life segment is better in itself for Abe than Abe*'s life segment for Abe*. The life segments may be internally

indiscernible, the only differentiating factor being that one contains a desire whose object obtains whereas the other contains a type- or near type-identical desire whose object does not obtain.[11] Similarly, in the cases involving Ken and Ken*, one may well be drawn to the view that authenticity (just like truth) does make a difference to well-being. Moderate "authenticity preferentists" may plausibly maintain that Ken*'s well-being is not enhanced (owing to the surreptitious finagling) as much as is Ken's.

What though, one may press, if manipulated Ken* now *endorses* the psychosurgery? Why is his well-being not now augmented as much as is Ken's well-being? If the "endorsement" is itself an outcome of the psychosurgery—if it is "engineered"—then I don't think that this should affect the verdict that Ken*'s well-being is not enhanced as much as Ken's. Suppose, though, that the endorsement is not the result of any finagling. Then, again, I don't see why the view that satisfaction of the relevant engineered-in desires should not contribute as much to Ken*'s well-being is mistaken. A comparison may help. If, unbeknownst to Ken*, his decision to aid the needy causally issues from an (unsheddable) engineered-in desire to aid the needy, he would not be morally responsible for this decision. Suppose he now discovers that he acquired this desire as a result of psychosurgery, and he now "endorses" this desire. Why should any of this throw into question the verdict that he is not responsible for his pertinent decision?

Second, if the satisfaction of inauthentic desires contributes to well-being, then we are forced to conclude that Beth's life in Matrix-4 is highly intrinsically good for Beth. But it is highly controversial that her life in the Matrix is highly good for her. Preferentists should, therefore, reject the view that satisfaction of inauthentic intrinsic desires contributes to well-being (or preferentists should, minimally, reject the view that satisfaction of inauthentic intrinsic desires contributes as much to well-being as does the satisfaction of authentic intrinsic desires). I grant, of course, that the second premise—that Beth's life in Matrix-4 is not highly good in itself for Beth—may well be disputed. But I think that the premise has a fair degree of plausibility. It should have this degree of plausibility especially for preferentists who want to resist purely mental state theories of welfare.

Third, suppose one is drawn to preferentist accounts of well-being. It is credible that, given that welfare is essentially personal, when we assess from a preferentist's perspective an agent's welfare, it is the agent's *own* desires or preferences and not the desires or preferences of someone else that are of concern. However, it is not quite so straightforward to specify what qualifies as one's *own* desires or preferences unless one pays close heed to normative authenticity. This requires elucidation.

Imagine that during a stretch of time manipulated Beth finds herself with the sorts of desire that Ann has; Beth would not have had these desires but for the manipulation. Suppose these engineered-in desires of Beth are satisfied. We ask again: does the satisfaction of these desires contribute to Beth's well-being? If we

---

[11] See, for example, Griffin (1986), Chapter 1; Sumner (1996, pp. 92–98); and Kagan (1998, pp. 35–36), for some pertinent literature on such cases.

answer in the affirmative, then why should it not also be the case that Beth's well-being is satisfied even when Beth lacks these desires but when Ann's desires of this sort are satisfied? The anticipated initial response would be that Beth simply does not *have* these desires. How can the satisfaction of *Ann's* desires contribute to enhancing *Beth's* well-being when Beth does not have such desires or does not have desires to the effect that Ann's desires be satisfied? This response presupposes a relatively simple view of desire ownership: one's desires are one's own if one has them.

But there is a problem with this easy response. Assume that the engineered-in desires are not of the sort that we find in the original Ann/Beth case. In the original case, the implanted desires, goals, and so on alter significantly the contours of Beth's life. Introducing a new character, assume that Betsy has fallen victim to the new-wave psycho-manipulators. The desires implanted in her are "copies" of some of the inconsequential desires of Annie. These desires ("IC-desires") have the following features: (a) they are not connected, in any obvious fashion, with the rest of the psychological constitution of the victim of manipulation (Betsy in our case). (b) They are disassociated from the victim's history. Betsy, for instance, has not acquired these desires as a result of practical deliberation or a normal upbringing. (c) They have nothing essential to do with the subject's goals, aims, cares, or concerns (whether or not the subject is a victim of manipulation). Finally, (d), the satisfaction of these desires does not contribute to the subject's enjoyment or suffering. In the case under consideration, each of Betsy's IC-desires is type- or near type-identical to the corresponding IC-desires of Annie. Suppose, despite Betsy's IC-desires instantiating features (a)–(d), and thus being highly impersonal and inconsequential, satisfaction of these (inauthentic) desires enhances Betsy's well-being. Although their satisfaction seems to make no apparent difference to how her life is going, one insists that their satisfaction contributes to Betsy's well-being. Then, it would seem, Betsy's well-being should also be satisfied when the "copies" of such IC-desires that Annie has are satisfied *even when* Betsy lacks such desires. Again, one might rejoin: how can *Betsy's* well-being be enhanced (on a preferentist's account) when *someone else's* IC-desires are satisfied? With IC-desires, though, this complaint rings hollow. Even when Betsy has these desires, they are dissociated from her history and other elements of psychology, they are not connected to her goals and cares, and their satisfaction does not contribute to her enjoyment or suffering.

Thus, it appears, if the satisfaction of (inauthentic) IC-desires of a subject who has these desires contributes to the subject's well-being on the preferentist account at issue, then the account should declare that the satisfaction of these sorts of desire should contribute to the well-being of the subject even when the subject lacks these desires as long as these sorts of IC-desire of *someone else* are satisfied. But surely the consequence of the proposition expressed by the prior sentence is one that preferentists should be highly reluctant to accept. So preferentism seems to find itself in a spot of trouble.

One way in which a preferentist may begin to address this concern is to concede the relevance of authenticity to well-being. We have noted that Betsy's IC-desires

are not authentic—they are not truly Betsy's own. It is open to the preferentist to claim that in virtue of not being authentic, the satisfaction of her IC-desires fails to enhance her welfare. On a preferentist view, well-being is augmented when one's *own* intrinsic desires are satisfied. On a relatively simple conceptualization of desire ownership, a desire is one's own if one has it. The Annie/Betsy quandary highlights the moral that the preferentist cannot settle on this simple notion of a desire's being one's own. A more robust notion is called for to overcome the problem involving the satisfaction of IC-desires. I suggest that if one is a preferentist, contrary to the objection at issue, one *should* concede that one's desires being or failing to be normatively wise authentic has a definite impact on well-being. If the agent's IC-desires are authentic, the agent is implicated (in a fashion in which we previously discussed in Chapter 5) in the acquisition of these desires. By virtue of such agent investment, it seems plausible to hold that the satisfaction of such desires should contribute to the agent's well-being but not to the well-being of another person who lacks these desires.

Summarizing the gist of my views concerning the relevance of authenticity to well-being if one is a preferentist, imagine that Able has certain intrinsic desires, they are authentic, and they are satisfied. A preferentist would not want to hold that Mable's well-being is enhanced by virtue of these desire satisfactions of Able. After all, these desires are *not* Mable's. Suppose, now, that Mable comes to possess these sorts of desire as a result of unsolicited "value engineering." There is a sense in which Mable does not have these desires: she has acquired them via processes that completely bypass her capacities of deliberative control; so they are not truly hers. Not being truly hers, they are somewhat like the desires of Able that she does not have: like Able's desires, these engineered-in desires do not "belong" to Mable. Hence, their satisfaction should not contribute as much to her well-being as should the satisfaction of otherwise similar desires that are authentic. The concern is not that Mable's implanted desires do not "fit," or "resonate with," or "complement" Mable's other desires. Rather, by virtue of how they are acquired, these desires are foreign to Mable's agency. Alternatively, imagine that Mable mistakenly believes that she has an intrinsic desire for something, and she believes that this desire is satisfied. These beliefs do not contribute to her welfare's being enhanced because she does not have the desire in the first place. Now suppose that she did acquire this desire but she did so as a result of unsolicited value engineering. This time, although she possesses the desire, it is not truly her own. I propose that satisfaction of this desire, because it is not truly Mable's own, should not contribute to welfare enhancement as much as it would have had it been authentic.

I concede that some people may reject the normative account of authenticity that I favor, or they may reject the proposal that this account is pertinent to welfare. Even so, I hope I have said enough to motivate the view that there is *an* authenticity requirement on well-being. When I say that there is an authenticity requirement on well-being, again, I intend to be understood as claiming that authenticity (or its lack) can *affect* well-being; satisfaction of inauthentic desires may contribute to well-being but not to the extent that they would have had they been authentic.

11.6 Source Incompatibilism and Well-Being

It remains to be shown that if some (plausible) version of freedom-sensitive preferentism is true, then freedom has a definite impact on well-being. Again, the influence of freedom on well-being can be perspicuously highlighted by assuming the source incompatibilist's position regarding the incompatibility of determinism and free action. According to the source incompatibilist, if all events are causally determined, one's having of a certain desire (an event) or the state in which one has a certain desire cannot be free because the event or state originates in sources beyond one's control. In other words, source incompatibilism implies that if determinism is true, then the having of none of our desires is free, where the freedom at issue is the freedom that moral responsibility requires.

Should, though, a source incompatibilist concede that if our desires are not free, they are *also* not authentic in the sense of "authentic" to which *freedom-sensitive preferentism* appeals? I have raised and addressed this sort of question in earlier parts of the book (Section 6.2). Here, my response is guided by reflection on some of Robert Kane's views on ultimate origination. Kane, a libertarian who endorses the principle of ultimate origination (*Principle O*), suggests that "underived origination or sole authorship" is desirable because such authorship is necessary for other things that human beings generally desire and are worth wanting (1996, p. 80). These include genuine creativity, autonomy or self-creation, desert for one's achievements, moral responsibility, being suitable objects of reactive attitudes, dignity or self-worth, a true sense of individuality or uniqueness as a person, life-hopes requiring an open future, freely given love and friendship, and its being true in the fullest sense that one acts of one's own free will (1996, p. 80). Kane adds that ultimate origination is also related to "objective worth." He develops the tale of Alan the artist to explain the notion of objective worth:

> Alan has been so despondent that a rich friend concocts a scheme to lift his spirits. The friend arranges to have Alan's paintings bought by confederates at the local art gallery under assumed names for $10,000 apiece. Alan mistakenly assumes his paintings are being recognized for their artistic merit by knowledgeable critics and collectors, and his spirits are lifted. Now let us imagine two possible worlds involving Alan. The first is the one just described, in which Alan thinks he is a great artist, and thinks he is being duly recognized as such, but really is not. The other imagined world is a similar one in which Alan has many of the same experiences, including the belief that he is a great artist. But in this second world he really is a great artist and really is being recognized as such; his rich friend is not merely deceiving him to lift his spirits. Finally, let us imagine that in both these worlds Alan dies happily, believing he is a great artist, though only in the second world was his belief correct (1996, p. 97).

Kane proposes that though Alan would feel equally happy in both worlds, many would concede that there is an important difference in the value in the two worlds for Alan. To concede this, for Kane, is to accept a notion of objective worth according to which the value in question is not just a function of subjective states or experiences. Kane comments, "I want to suggest that the notion of ultimate responsibility is of a piece with this notion of objective worth" (1996, p. 97).

The point to which I am building up is the following. Should an incompatibilist believe that freedom in the sense of "freedom" which implicates our being ultimate originators is necessary for goods such as autonomy and objective worth, it seems highly credible that such an incompatibilist should also agree that this sort of freedom is required for welfare, provided that an appropriate version of freedom-sensitive preferentism is true. The reasoning for this claim is straightforward: if our having of desires originates in the distant past and by the natural laws, the having of these desires (according to source incompatibilists who accept *Principle O*) is not free; if the desires are not free because we are not their ultimate originators, the desires are not authentic; they are not "truly our own." Hence, on robust versions of freedom-sensitive preferentism, such desires cannot contribute to well-being. It follows, given the source incompatibilist's understanding of freedom, that if a robust version of freedom-sensitive preferentism is true, our lives are shorn of intrinsic value.

Some might see this drastic conclusion as a *reductio* of robust freedom-sensitive preferentism. So they may opt for a moderate version. On moderate versions, the satisfaction of inauthentic desires contributes less to well-being than the satisfaction of otherwise similar desires that are authentic, and the frustration of inauthentic desires contributes less to the intrinsic badness of a life than the frustration of otherwise analogous desires that are authentic. If such a moderate version is true, then once again freedom or its absence has a definite impact on welfare. Reflect on three types of life, each lived out in a determined world. In the first, the agent's life contains a far greater balance of intrinsic desire satisfaction than intrinsic desire frustration. We can say that none of these desire satisfactions or frustrations is free because none of the agent's desires is authentic (if one accepts pertinent source incompatibilist presuppositions). If the moderate version of freedom-sensitive preferentism under scrutiny is true, this agent's life will be less good for him than it would be if the corresponding desires—type- or near type-identical desires—were authentic. In the second, the agent's life contains a far greater balance of intrinsic desire frustration than intrinsic desire satisfaction. The moderate version implies that this life is better for the agent than a life in which the corresponding desires are authentic. Finally, in the third, the agent's life contains a roughly equal balance of intrinsic desire satisfactions and intrinsic desire frustrations. Maybe such a life would be just as good for its agent as it would be if the corresponding desires were authentic, assuming that the moderate version of freedom-sensitive preferentism were true.

Wrapping up, welfare is essentially personal. Bykvist (though addressing concerns somewhat different than those addressed in this chapter) gives powerful expression to this sentiment:

> To successfully pursue one's life-time ambitions, and to live in accordance with one's overall plan of life are ways to improve one's well-being. This intuition need not be based on any perfectionist ideas; it can be based on the compelling idea that a life full of meaning and success is a prudentially good life. Whereas our aims give our lives meaning and purpose, their accomplishments make our lives successful. . . . What is important. . .is that aims and ambitions are desires that concern the way something is brought about. To aim at something is to desire to do something or take part in the bringing about of some state of affairs. My

> aim cannot be identified with a desire whose content is that something is the case, e.g., that a book is written. My aim must be identified with a desire whose content is that something is brought about by me, e.g., that a book is written *by me* (2002, p. 483).

I have proposed that something like what Bykvist suggests is true of aims is more or less true of desires whose satisfactions or frustrations a preferentist should think contribute to well-being: these desires must be *mine*; they must be authentic. Reflect, again, on what attracts many people to the preferentist's general view of well-being: often enough, our desires express aims that we deem significant, and we think that fulfillment of these aims contributes to our well-being. If achievement of these aims is to enhance our well-being, surely these aims must be *ours*; they can't be aims imposed or implanted in us from "the outside." Thus, preferentists should welcome an authenticity requirement on well-being. If a version of preferentism (robust or moderate) that incorporates such a requirement is true, then there is a significant connection between welfare, on the one hand, and freedom, on the other.

Chapter 12
Freedom and Whole-Life Satisfaction Theories of Welfare

12.1 Introduction

In his insightful book, *Welfare, Happiness, and Ethics*, Sumner defends a certain view about happiness—a whole-life satisfaction view. He then develops an account of welfare—what he calls "welfare as authentic happiness"—which invokes this view of happiness. It is Sumner's theory of welfare that is the topic of this chapter. Sumner's theory is engaging for several reasons. First, it is of intrinsic interest; in my mind, it is the most well-developed and perceptive whole-life satisfaction theory in the pertinent literature. Second, during the course of refining his theory, Sumner says a number of penetrating things about well-being in general. Third, Sumner's theory *can* be construed as a version of a freedom-sensitive axiology. If a theory of the sort that Sumner defends is true, then, as I explain, hard incompatibilism has a pronounced effect on welfare.

In what follows in this chapter, I first reconstruct what I hope is correctly representative of Sumner's theory. Then I advance some critical comments against the theory. Finally, I clarify why hard incompatibilism imperils personal well-being if Sumner's theory is true.

12.2 An Outline of Sumner's Theory

We begin with salient aspects of Sumner's whole-life satisfaction (SWLS) account of happiness. In the context of interest, happiness refers to some psychological state. To say that a person is happy (in the relevant sense) is simply to describe him. In itself, it is not a prudential, or for that matter, a moral or aretaic evaluation. A central project in the happiness literature is to discover the nature of happiness. Ideally, what is sought is a statement of the necessary and sufficient conditions for a person's being happy. (Another project that should not be conflated with the first is to catalogue the likely sources of happiness.) Philosophical theories about the nature of happiness fall into some main families. One of them is hedonism (see, e.g., Feldman n.d.). Another seems to make happiness turn on flourishing, or virtue.[1] A third is the

[1] Some people claim to find this view in the writings of Aristotle.

I. Haji, *Freedom and Value*, Library of Ethics and Applied Philosophy 21,

preferentist family. Theories in this family make happiness fundamentally a matter of the satisfaction of desires (see, e.g., Davis 1981a,b; and Kahneman 1999). Whole-life satisfactionism may be construed as a version of preferentism about happiness.[2] According to theories of this sort, to be happy is to be satisfied with your life as a whole.

A number of distinguished philosophers have defended whole-life satisfaction accounts including, in addition to Sumner, Richard Brandt (1967, 1979, 1989), Robert Nozick (1989), John Kekes (1982, 1992), Elizabeth Telfer (1980), and Wladyslaw Tatarkiewicz (1976). Brandt, for example, writes:

> [T]he following proposal for a definition of "happy" may be suggested. . . .[I]n order to be happy it is necessary that one like. . .those parts of one's total life pattern and circumstances that one thinks are important. To say that one likes them is in part to say that one is "satisfied" with them—that one does not wish them to be substantially different, and that they measure up, at least roughly, to the life ideal one had hoped to attain....[We] would not call a man happy if he did not frequently feel joy or enthusiasm or enjoy what he was doing or experiencing (1967, pp. 413–414).

Tatarkiewicz says that happiness is "satisfaction with one's life as a whole" (1966, p. 1).

> Satisfaction with life as a whole must be satisfaction not only with that which is, but also with that which was and that which will be, not only with the present, but also with the past and the future (1976, p. 140).

Tefler explains:

> My suggestion for a definition of happiness, then, is that it is a state of being pleased with one's life as a whole . . . (1980, pp. 8–9).
>
> [A] happy man does not want anything major in his life to be otherwise; he is pleased with, that is wants (to keep), what he has got; there is nothing major which he has not got and which he wants (to get) (1980, p. 8).

Toward developing his whole-life satisfactionism, Sumner cautions that "happiness" and "unhappiness" are ambiguous. He distinguishes "four kinds or dimensions" of happiness:

1. *Being happy with or about something*. Sumner says that a distinctive feature of this kind of happiness is that

   > it requires completion by an intentional object: there must be something with or about which you are happy. Any state of the world can serve as such an object as long as you have a positive attitude toward it, or regard it favorably, or like or approve of it. The favorable attitude need not be a strong one. Being happy with something is roughly equivalent to being satisfied or content with it, i.e., finding that it measures up reasonably well to some standard . . . . Contrariwise, being unhappy with something means regarding it in an unfavorable light, though not as an absolute disaster (1996, pp. 143–144).

---

[2] The views of Brandt (1967, 1979, 1989), Telfer (1980), Nozick (1989), Sumner (1996), and Tatarkiewicz (1976) fall into this category.

2. *Feeling happy.* This contrasts with the first sort of happiness in two primary respects: it involves an occurrent feeling and requires no intentional object. "Not quite grand enough to count as an emotion, it is rather a mood of optimism or cheer which colours your outlook on your life and on the world in general, concentrating attention on everything that is positive and upbeat" (p. 144). Sumner clams that the negative counterpart to this mood is feeling unhappy, down, dejected, or depressed. He adds that elation and depression are the opposite poles of a continuum, which admits of many intermediate stops—"you may be up but not euphoric, down but not suicidal" (p. 144).
3. *Having a happy disposition/personality*. Feeling happy (unhappy) is an occurrent disposition in your life. You may have a settled tendency toward such positive (negative) moods. If you do, you have a happy disposition or you are a happy person.
4. *Being happy/having a happy life*. It is this notion to which Sumner's account of welfare appeals. Sumner says:

> Being happy in this sense means having a certain kind of positive attitude toward your life, which in its fullest form has both a cognitive and affective component. The cognitive aspect of happiness consists in a positive evaluation of the conditions of your life, a judgement that, at least on balance, it measures up favourably against your standards or expectations. This evaluation may be global, covering all of the important sectors of your life, or it may focus on one in particular (your work, say, or your family). In either case it represents an affirmation or endorsement of (some or all of) the conditions or circumstances of your life, a judgement that, on balance and taking everything into account, your life is going well for you.
>
> Clearly this sort of prudential stocktaking is possible only for creatures capable of assessing their lives as wholes, either at a time or over some extended period of time. The cognitive component of happiness is, therefore, beyond the range of many subjects-of-a-life, such as small children and nonhuman animals. However, there is more involved in being happy than being disposed to think that your life is going (or has gone) well. The affective side of happiness consists in what we commonly call a sense of well-being: finding your life enriching or rewarding, or feeling satisfied or fulfilled by it. Because it is less cognitively demanding than a judgment about how one's life is going as a whole, it is what we have in mind when we say that a child or an animal is happy, or is leading a happy life (pp. 145–146).

Analogously, being unhappy is a matter both of evaluating your life (or some part of it) negatively, as failing to meet your standards for it, and of experiencing it as unsatisfying or unfulfilling (p. 146, n. 4).

Suppose a person is happy to a certain degree, n, at a time or over an interval of time. Sumner proposes that the best way to determine the degree to which a person is happy is to ask that person, and he claims that social scientists have developed a sophisticated methodology for eliciting self-assessments of life satisfaction (p. 172). He stresses that an individual's self-assessments are reliable only if they are relevant, sincere, and considered. Relevance concerns getting the question about the sort of assessment at issue right (p. 154). Subjects must assess the *prudential value* of their lives: they must judge whether their lives are going well for *them* or not and, for instance, whether their lives are aesthetically or morally good (pp. 153–154). The judgments concerning the prudential value of a life are sincere when they are frank

and honest; they are uninfluenced, for instance, by a desire to maintain a certain social image (pp. 154–155). These judgments are considered when they are not clouded by transitory moods (p. 155).

Moving, now, to Sumner's account of welfare, suppose that a person judges that her life at time *t* or some interval is going well for her (to degree *n*), by her standards for her life; suppose her judgment is relevant, sincere, and considered; and suppose she experiences her life as fulfilling. Would it then be correct to identify her degree of happiness with her degree of well-being? Sumner resists this identification for two reasons. A happiness judgment of *S* is a relevant, sincere, and considered judgment that *S* makes regarding how well, on balance, *S*'s life measures up against *S*'s life ideal. First, Sumner says that such judgments may not be adequately informed and so may involve errors (pp. 156, 162, 167, 174, 175). He recommends that a subject's happiness contributes toward well-being only if her judgment regarding her happiness level is not factually uninformed or misinformed. He proposes that if a happiness judgment involves such errors, it will not affect the degree of *happiness* but the degree of *welfare* (p. 158). Reconsider the deceived businessman. As long as he does not discover the deception, he may well be happy, but according to Sumner, it seems, his welfare level won't coincide with his happiness level.

This problem concerning the possibility of factual error is rectified, in part, by imposing an information constraint on happiness judgments. The happiness judgments, if they are to qualify as welfare judgments, must meet an information requirement. Very broadly, the requirement says that subjects must be appropriately informed about the conditions of their lives. But when are they appropriately informed? Must the factual information which informs the judgments be true, for example, or justified if the subjects are to be aptly informed? Addressing this issue, Sumner writes:

> [W]hich epistemic conditions are appropriate? The strongest candidate would be a truth or reality requirement, which would stipulate that happiness counts as well-being only when it is based on a view of the conditions of our lives which is free from factual error. . . .[T]his stipulation would be unreasonably puritanical. We do not invariably reassess earlier periods of happiness in this austere manner once we realize the extent to which they depended on false beliefs about states of the world: the intentions of a lover, the integrity of a public figure, the prospects of success at a new enterprise, or whatever. . . . When we reassess our lives in retrospect, and from a superior epistemic vantage point, *there is no right answer to the question of what our reaction should be*—that is surely up to us. Because a reality requirement stipulates a right answer—any happiness based on illusion can make no intrinsic contribution to our well-being—it must be rejected as presumptuously dogmatic (pp. 158–159).

Rather than calling for an information requirement that demands that a person's information that bears on her happiness judgments be true or justified, Sumner proposes that to qualify as welfare judgments, her happiness judgments must satisfy a defeasibility constraint: they are "authoritative" unless there is evidence that they are not informed—there is evidence that either they are factually uninformed or misinformed. We may now formulate the "information requirement" in the following way:

Information Requirement: S's relevant, sincere, and considered judgment that, on balance, *S*'s life measures up to degree *n* against S's life ideal, is to be taken at face

value (or, alternatively, correctly reflects the fact that *S*'s life measures up to degree *n* against *S*'s life ideal) only if there is no evidence that this judgment is factually informed or misinformed.

Tying some ends together, Sumner recommends:

S's life is going well for *S* only if *S* judges that, on balance, *S*'s life measures up favorably against *S*'s own standards (or expectations) concerning a good life (pp. 153–155, 162).

Abridging somewhat:

S's life is going well for *S* only if S judges—*S* makes a happiness judgment—that, on balance, *S*'s life measures up favorably against *S*'s life ideal.

A happiness judgment is informed if and only if it satisfies the *information requirement*. So we can say:

S's life is going well for *S* only if *S* makes an informed happiness judgment.

This amounts to the following:

S's life is going well for *S* only if (i) *S* judges—*S* makes a happiness judgment—that, on balance, *S*'s life measures up favorably against *S*'s life ideal; and (ii) this judgment of *S* is relevant, sincere, considered, and informed.

Recall the affective component of happiness: to be happy, one must feel satisfied or fulfilled with one's life. The feeling dimension of happiness that we previously recorded supplies the pertinent interpretation of "feeling satisfied with one's life." Including the affective component in the partial analysis yields the following:

S's life is going well for *S* only if (i) *S* judges—*S* makes a happiness judgment—that, on balance, *S*'s life measures up favorably against *S*'s life ideal; (ii) this judgment of *S* is relevant, sincere, considered, and informed, and (iii) *S* feels satisfied or fulfilled with *S*'s life.

Sumner rejects the identification of "mere" happiness with welfare for a second reason. (He rejects, i.e., the view that *S* is happy to degree *n* at *t* if and only if *S*'s life is good in itself for *S* to degree *n* at *t*.) Indeed, he claims that factual error is not the only or most important reason for questioning the accuracy of happiness judgments. Another factor is that the judgments may be nonautonomous. They may be derivatively inauthentic in that they may be based—they causally derive from—values, beliefs, deliberative principles, and so forth that are not "truly the agent's own" or not authentic. Sumner writes:

> A person is autonomous when her beliefs, or values, or aims, or decisions, or actions are, in some important sense, *her own*. There is therefore an evident connection between autonomy and what we have been calling authenticity. We have said that a subject's affirmation or endorsement of her life is made from her own point of view, is truly *hers*, only when it is authentic. The demand that self-evaluations be authentic flows from the logic of a subjective theory, which grounds an individual's well-being on *her* (positive and negative) attitudes.

> One side of this demand is the requirement that subjects be informed about the conditions of their lives. The other is the requirement that they be autonomous (1996, p. 167).

To motivate an autonomy constraint on Welfare, Sumner cites Amartya Sen's examples of the hopeless beggar, the precarious landless laborer, the dominated housewife, the hardened unemployed, and the overexhausted coolie (1996, p. 162):

> A person who has had a life of misfortune, with very little opportunities, and rather little hope, may be more easily reconciled to deprivations than others reared in more fortunate and affluent circumstances. The metric of happiness may, therefore, distort the extent of deprivation, in a specific and biased way. The hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie may all take pleasures in small mercies, and manage to suppress intense suffering for the necessity of continuing survival, but it would be ethically deeply mistaken to attach a correspondingly small value to the loss of their well-being because of this survival strategy (1987, pp. 45–46).

Sumner explains that we are reluctant to take at the face value the life satisfaction reported by these people because "we suspect that the standards which their self-assessments reflect have been artificially lowered or distorted by processes of indoctrination or exploitation" (p. 166). The remedy is to correct for the conditions under which their expectations about themselves came to be formed: "The problem is not that their values are objectively mistaken but that they have never had the opportunity to form their own values at all. They do not lack enlightenment, or insight into the Platonic form of the good; they lack autonomy" (p. 166).

Sumner submits that authenticity is a core element in the concept of autonomy, "just as subject-relativity is a core element in the concept of welfare" (p. 167). Though he does not develop an account of authenticity, he proposes the following.[3]

> However the details of a fully adequate view [of authenticity] are worked out in the end, the implications for our theory of welfare are clear. Self-assessments of happiness or life satisfaction are suspect (as measures of well-being) when there is good reason to suspect that they have been influenced by autonomy-subverting mechanisms of social conditioning, such as indoctrination, programming, brainwashing, role scripting, and the like. Since these are all socialization processes, and since we are all historically embedded selves, the practical question becomes how much emancipation from her background and social conditions a subject must exhibit in order for her self-assessment to be taken at face value. As in the case of the information requirement, the best strategy here is to treat subjects' reports of their level of life satisfaction as defeasible—that is, as authoritative unless there is evidence that they are non-autonomous (1996, p. 171).

We may now formulate the "autonomy requirement":

Autonomy Requirement: *S*'s relevant, sincere, and considered judgment that, on balance, *S*'s life measures up to degree *n* against *S*'s life ideal, is to be taken at face value (or, alternatively, correctly reflects the fact that *S*'s life measures up to degree *n* against *S*'s life ideal) only if there is no evidence that this judgment is derivatively

---

[3] He *does* discuss a couple of views, Frankfurt's hierarchical view (1971), and John Christman's historical view (1991).

inauthentic. (Such a judgment would be derivatively authentic if it causally arose from desires, beliefs, values, and the like that are inauthentic.)

Now we may say that a happiness judgment is autonomous if and only if it satisfies the *autonomy requirement*. Finally, we can summarize what may be Sumner's whole life satisfaction account of welfare (SWLS) in this way:

SWLS: *S*'s life is going well for *S* if and only if (i) *S* judges—*S* makes a happiness judgment—that, on balance, *S*'s life measures up favorably against *S*'s life ideal; (ii) this judgment of *S* is relevant, sincere, considered, informed, and autonomous, and (iii) *S* feels satisfied or fulfilled with *S*'s life.

12.3 Some Comments on Sumner's Theory

Sumner's whole life satisfactionism regarding welfare merits serious consideration. But this rich account raises a number of questions as well.

We note, initially, that if taking pleasure in "base objects" is a concern for simple intrinsic attitudinal hedonism (and, indeed for simple preferentism, too), then it may well also be a concern for SWLS. Simply imagine that Porky's ideal of the good life (for him) is that he wallows with the pigs. Suppose that at various times during his life, Porky makes happiness judgments to the effect that, on balance, his life measures up favorably against his life ideal. I see little reason to deny that his judgments may well be relevant, sincere, considered, informed, and autonomous. Let's add that Porky feels highly satisfied with his life. Then SWLS, in conjunction with these facts, implies that Porky's life is good in itself for Porky.

One potential problem with SWLS is what we might call the "scope" problem. In a passage cited above, Sumner submits that the sort of "prudential stocktaking" that the cognitive component of *being happy* involves "is possible only for creatures capable of assessing their lives as wholes, either at a time or over some extended period of time. The cognitive component of happiness is therefore beyond the range of many subjects-of-a-life, such as small children and non-human animals" (pp. 145–146). Sumner concedes, though, that unlike the cognitive component, the affective component of happiness need not be beyond the reach of such creatures.

Suppose Tanto is five. His life has been a happy one. He has never, though, engaged in the relevant sort of prudential stocktaking concerning his life—he has never made a happiness judgment—because he does not have the ability to do so. If SWLS is true, it seems that he gets a score of 0 on SWLS's account of well-being. This account implies that his life has never been good in itself for him. This is, in part, because on Sumner's whole life satisfactionism concerning *happiness*, it appears that Tanto has never been happy. If this is indeed a result of the theory (in conjunction with appropriate facts of the case), the result is unfortunate.

Introducing Zen (an adolescent), at a particular time, *t*, Zen is injured in an accident. Prior to *t*, Zen like Tanto, had never reflected on his life as a whole not because he lacked the cognitive abilities to do so—he did have these abilities over stretches of his life—but simply because he had never thought about this deep issue. The

injury at *t* renders him incapable of engaging in such whole-life reflection. Still, he can enjoy many things—he takes intrinsic pleasure in many different things—and he can suffer as well—he takes intrinsic displeasure in other things. Once again, SWLS generates a remarkable result: it implies that there is no time at which Zen's life has been good in itself for Zen owing, in part, to Zen's never having been happy (in the welfare-pertinent sense of "being happy" at issue) in his life. Again, this seems erroneous.

Sumner may rejoin that both Tanto and Zen *have* been happy. He explains:

> For the happiness theory, the minimal wherewithal for having a welfare is being a subject who is capable of being satisfied or unsatisfied by the conditions of one's life. In the case of paradigm human subjects with complex cognitive capacities, more is necessary as well: their judgements about the quality of their lives must be authentic. Where these more sophisticated skills are absent, the *sine qua non* is the baseline ability to experience one's life, in the living of it, as agreeable or disagreeable. The most primitive form of this ability is the capacity for enjoyment and suffering, or for pleasure and pain. If we call this capacity *sentience*, then we may say that on the happiness theory the class of core welfare subjects is populated by all sentient creatures (1996, p. 178, note omitted).

Sumner's view seems to be that since both Tanto and Zen satisfy the affective requirement of happiness, this is sufficient for the two being happy. And perhaps, though Sumner does *not* commit himself to this, as the two *are* happy, there is no need to deny that their lives have been good in themselves for them.

This response, however, generates further concerns. First, imagine that Ken is in his early fifties. So far his life has been a happy one. Assume that he judges that, on balance, his life is going well for him. His judgment is sincere, considered, informed, and autonomous, and he feels satisfied with his life as a whole. Suppose Kenzo is five. His life, too, has been a happy one. Like Tanto but unlike Ken, Kenzo has never engaged in prudential whole-life stocktaking. He does, though, feel satisfied with his life; thus, he fulfills the affective component of *being happy*. When we say "Ken's life is happy" and "Kenzo's life is happy," presumably, we use "happy" in the same sense in each of these utterances. Alternatively, "happy" expresses the same notion of *being happy* in each of the utterances. Sumner's rejoinder implies otherwise. The passage that I last quoted suggests that there is a "thick" notion of happiness that has both a cognitive and affective component and a "thin" notion that has only an affective component. In the first utterance, "Ken's life is happy," "happy" expresses the strong notion; in the second utterance, "Kenzo's life is happy," "happy" expresses the thin notion. I, for one, have strong prima facie doubts whether we are using "happy" in two different senses in utterances such as these.

Second, if we assume (contrary to what SWLS implies) that both Tanto's and Zen's life *are* good in themselves for them, and they are so *because* they contain a net balance of enjoyment (intrinsic attitudinal pleasure), then it is unclear why an apt version of attitudinal hedonism should not claim victory over SWLS. It would be theoretically out of the ordinary to submit that Ken's life is SWLS-good-in-itself for Ken but that Zen's life, though not good in itself for Zen in this fashion, is, say, SIAH-good-in-itself for Zen.

There is a second problem with SWLS. Remy subscribes to a life ideal that it is not possible for him to attain. He judges, at a certain time, that his life falls far short of satisfying this ideal. His judgment is relevant, sincere, considered, informed, and autonomous. Reflecting on the fact that he cannot hope to attain his ideal, he ceases to care about attaining it and decides to enjoy life to the fullest. He takes intrinsic pleasure in many and varied things; sometimes he suffers intrinsic displeasures. Further, on the whole he feels happy with his life, all the while (at least dispositionally) not losing sight of the fact that his life does not measure up to his life ideal. SWLS implies that his life is not good in itself for him. But, overall, it seems that he is a happy guy who enjoys a goodly measure of well-being.

There seems yet to be another concern with SWLS. To understand this concern, we start with a passage (part of which has been previously quoted) in which Sumner summarizes his reasons for rejecting any hedonistic theory of welfare which "treats pleasure and pain simply as mental states" (p. 98):

> Any such theory will entail that the impact on our well-being of some particular experience is entirely determined by features of the experience which are available to introspection—how it feels, how agreeable we find it, how much we wish it to continue, or whatever. We may therefore track how well our lives are going just by attending to how they seem from the inside, bracketing off all questions of their anchoring in the external world. The lesson of the experience machine is that any theory with this implication is too interior and solipsistic to provide a descriptively adequate account of the nature of welfare. Since welfare does not consist merely of states of mind, it does not consist merely of pleasurable states of mind, regardless of how these are characterized (1996, p. 98).

Reflect, yet again, on the case of the businessman or Neo in Matix-1. Even the simple theory (just like default sensory hedonism) implies, unacceptably it would seem as far as Sumner is concerned, that the deceived life is good in itself for the businessman and that the fabricated life is good in itself for Neo. Sumner's happiness theory is supposed to evade this objection from false pleasures in virtue of its demand that subjects of welfare satisfy the *information* and *authenticity* requirements:

> [Distinguish] two types of subjective theory: those on which welfare is *solely* a matter of the subject's states of mind and those on which it is *additionally* a matter of some states of the world. Classical hedonism is a mental state theory, vulnerable to objections from illusion or deception for this very reason. The desire theory is a state-of-the-world theory, since whether or not a desire is satisfied is a matter of how the world goes. In some important respects the happiness theory resembles hedonism, most obviously in its endorsement of an experience requirement: according to the theory, it is a necessary condition of a state of affairs making me better off (directly or intrinsically) that it enter into my experience. It was the absence of an experience requirement which was fatal to the desire theory. However, in an equally important way the happiness theory sides with the desire theory against classical hedonism. Since it also incorporates an information requirement (as part of its condition of authenticity), it is a state-of-the-world theory. That I experience a state of affairs is necessary in order for it to benefit me, but (since the experience may be illusory or deceptive) it is not sufficient (1996, pp. 174–175).

In a revised story of the deceived businessman, suppose the businessman has strong suspicions that he is being deceived about his relationships, his social status,

and other matters. But suppose he correctly judges that, all things considered, he would be better off if he were to mask the deception from himself and continue living as he does. Suppose his pertinent endeavors of self-deception are successful. If, in the original case, one judges that the businessman's life is not good in itself for him, then, it seems, one should judge as well that in the revised case, his life is not good in itself for him. In both cases, the businessman is deceived about the circumstances of his life. In the second case, unlike in the first, he suspects deception but succeeds in masking the deception from himself. In either case, the two lives "feel" the same from the "inside."

Does the *information requirement* shield the happiness theory from yielding an analogous result in the self-deception case, a result that Sumner would, presumably, want to shun? Commenting on this requirement, Sumner says:

> The problem with reality or justification requirements is that they impose uniform discount rates on everyone alike: happiness has no prudential payoff unless fully informed, or is discounted at a steady rate as it becomes less informed. The relevance of information for a person's well-being is a personal matter to be decided by personal priorities; there is here no authoritative public standard. Still, the problem remains that the self-assessments which individuals report cannot merely be taken at face value; we need to know whether they are authentic. The best way to capture the condition they must satisfy is to say that they are *defeasible*—that is, they are authoritative unless we have some reason to think that they do not reflect the individual's own deepest priorities. Where someone's endorsement of his life is factually uninformed or misinformed, that gives us one reason for doubting its authority (whether it is a sufficient reason depends on whether the endorsement will, or would, survive the acquisition of the missing information) (1996, p. 161).

In the revised story, there is little reason to think that living a deceived life does not reflect the businessman's own deepest priorities. Additionally, the businessman's endorsement of his deceived life survives the acquisition of the information that he is being deceived. Thus, one might initially conclude that both the happiness theory and, for instance, the simple theory generate the result that (in the revised story) the businessman's life is good in itself for the businessman. Then one may go on to argue that if you are inclined to reject the simple theory because of the results it generates in the original case, there is little reason why you should not also be inclined to reject this theory on the basis of the results that it yields in the revised case.[4] Further, if you reject the simple theory in virtue of its relevant implications concerning the revised case, you should also reject the happiness theory because of its analogous implications.

However, one may attempt to resist this entire line of reasoning for the view that the happiness theory generates the same results as does the simple theory in the revised case involving self-deception. For suppose the self-deceived businessman forms a happiness judgment about his life. In the throes of self-deception, his happiness judgment seems *not* to be sincere. Regarding sincerity, Sumner claims that a subject's prudential stocktaking

[4] Maybe the proponent of the simple theory would claim that in the revised case, the deceived businessman is happy but his welfare level is low.

> may fail to reflect his true state of mind because it is *insincere*. He may, for instance, be influenced by a (conscious or unconscious) desire to measure up to some self-imposed standard or some (perceived) expectations on the part of others, as a result of which he tells us not how happy he actually is but how happy he thinks (or he thinks we think) he ought to be. There is considerable experimental evidence which suggests that subjects systematically overstate their levels of life satisfaction, so as to represent themselves to others (and perhaps also to themselves) as happier than they actually are (1996, p. 154, note omitted).

If the self-deceived businessman's happiness judgment is not sincere, then SWLS implies that this judgment does not accurately reflect his welfare level. Thus, it might be rejoined, SWLS does *not* have the disturbing implication that one might initially think it has in the self-deception case—that the deceived businessman's life is just as good for him as is the life of the undeceived businessman.

If one is inclined to maneuver in this way to evade the result that the deceived businessman's life is good in itself for him, then there is another worry. It seems intuitively plausible that while in the state of self-deception, a person may enjoy a high level of well-being. The maneuver in question appears to imply otherwise: the happiness judgments of the self-deceived person are not sincere (or, perhaps, they are not informed). This, in turn, casts suspicion on whether the sincerity requirement on happiness judgments (or perhaps, the information requirement, in general) is too stringent.

12.4 Hard Incompatibilism, Authenticity, and Whole-Life Satisfactionism

What bearing, if any, does hard incompatibilism have on well-being if SWLS is true? As I have formulated it, SWLS requires that the happiness judgment that the subject makes if she is to enjoy well-being must be derivatively authentic; it does not imply that the subject's relevant feeling constitutive of the affective component of happiness be authentic. It would seem, though, that if one insists on the authenticity of the cognitive component of the happiness theory, one should have little reason not to insist on the authenticity of its affective component as well: the subject's feelings concerning her whole life must be (derivatively) authentic. The third clause of SWLS should thus be amended along these lines:

(iii-Revised): *S* feels satisfied or fulfilled with *S*'s life and these feelings or affective states of *S* are (derivatively) authentic.

It may be worth stressing that the account of authenticity on which SWLS draws shoulders a considerable burden. Suppose, though it should be evident that I do not agree with this, that some hierarchical account of authenticity is correct. On one incarnation of this view, a first-order desire, for instance, is truly the agent's own if and only if its agent has an unopposed second-order desire that this first-order desire moves the agent all the way to action. It is quite possible that on such a view of authenticity, even manipulated Beth's engineered-in desires turn out to be authentic. I believe that this is a result that Sumner would not find congenial. I

do not wish to imply that Sumner is committed to any such hierarchical view of authenticity. Indeed, he seems to reject this sort of view (1996, pp. 168–169). I simply wish to highlight the fact that whether the *authenticity* and *information* requirements can do the work that Sumner wants them to do—for example, blocking concerns of deception that Sumner would judge plague default sensory hedonism or the simple theory—depends crucially on which account of authenticity wins the day.

Assume that SWLS or some close kin is in the right ballpark. Would freedom or lack of it affect life-ranking axiological judgments? I have proposed that if a feeling-token (or state) is authentic, its authenticity derives from the authenticity of decisions. I have also argued that the hard incompatibilist is committed to the view that in hard incompatibilist worlds (worlds devoid of agent causation that are either deterministic or indeterministic), no springs of action are authentic and, thus, no decisions are derivatively authentic (see Section 5.2). This, in turn, implies that no feeling-tokens or states in such worlds are authentic. It would follow that if hard incompatibilism is true, no subject in any hard incompatibilist world could satisfy clause (iii-revised) of SWLS and, consequently, no life for any subject in such worlds would be good in itself for that subject if SWLS were true. The combination of hard incompatibilism and the happiness theory, hence, has some drastic implications for well-being.

12.5 Overall Conclusion

To wrap up briefly, it is an interesting project to uncover the freedom presuppositions (if any) of various species of moral judgment. We are, for instance, familiar with an age-old, fundamental question in the free-will debate: is the freedom that moral responsibility requires—is free will—compatible with determinism? Incompatibilists have championed the thesis that determinism undermines the truth of judgments of moral blameworthiness and moral praiseworthiness; compatibilists have denied this thesis. We are also familiar, though perhaps to a lesser degree, with the question of whether freedom or its lack influences the truth of obligation judgments. If freedom and so, perhaps, determinism have an impact on the truth of responsibility judgments and obligation judgments, it is only natural to wonder whether freedom and so again, perhaps, determinism have an impact on the truth of axiological appraisals having to do with welfare value and worldly value. This theoretical issue has been given relatively little consideration even though it is philosophically significant: first, it is of intrinsic interest. A fundamental project in moral philosophy is getting clear on elements that contribute, either positively or negatively, to personal well-being. Suppose incompatibilists in the free-will controversy have won the day—determinism undermines free will. Suppose, further, that it is also true that the atoms that contribute to the intrinsic value of lives and worlds, if free, are better than otherwise similar unfree atoms. Then determinism will have a *negative* effect on personal well-being and worldly value. Second, as we have registered, on a highly prominent view of what our moral obligations consist in, we ought to do what we do in the intrinsically best worlds accessible to us. If determinism has a

detrimental impact on the value of worlds, then on this prominent view, determinism has an impact on moral obligation. Third, assume that it can be shown, as I have argued, that the freedom of the atoms, on influential, leading theories about welfare value or worldly value, traces to, or is derivative from the freedom of various mental actions, such as making decisions. Suppose, for example, that taking intrinsic pleasure in something, if free, is indirectly free in that the freedom of such an episode of pleasure derives from the freedom of things—mental actions in my estimation—with respect to which we are (directly) free. Then this exposes another route for the incompatibilist to argue for the view that determinism is more menacing than it has hitherto been appreciated to be: owing to determinism's undermining the freedom of our decisions, determinism has a definite impact on various, significant appraisals regarding worldly value and personal well-being.

I adopted the following theoretical framework to explore the freedom presuppositions of life- and world-ranking axiological judgments. First, I inquired into whether free "positive" atoms—freely taking intrinsic pleasure in things, for example—are intrinsically better than otherwise similar atoms that are unfree, and whether free "negative" atoms—freely taking intrinsic displeasure in things, for instance—are not as bad as otherwise similar unfree atoms. This is in keeping with the strategy that other axiologists have adopted to argue, for instance, for the view that truth contributes to personal well-being—taking pleasure in true states of affairs is better, for example, than taking pleasure in false states of affairs (e.g., Kagan 1998, Adams 1999); or that object-worthiness contributes to personal well-being—taking pleasure in objects that deserve to have pleasure taken in them is better, for example, than taking pleasure in objects that do not deserve to have pleasure taken in them (e.g., Moore [1903] 1962, Chapter III, Section 56; Feldman 2004).

I suggested that this first strategy can be duly pursued only in association with substantive views about the *good life* or substantive views about the intrinsic value of worlds. In large measure, the substantive theories to which I restricted attention are forms of hedonism. But I also discussed some varieties of preferentism and a rich version of a whole-life satisfaction account of welfare.

Second, having made the initial case that free positive atoms (on the substantive views at issue) are better either for welfare or worlds than otherwise similar atoms that are not free, and that free negative atoms are not as bad, again either for welfare or worlds, as otherwise similar unfree atoms, I scrutinized the elements of these atoms. I asked whether these elements themselves have freedom presuppositions. We noted, for example, that many have proposed that deserved intrinsic pleasures are better than corresponding pleasures that are not deserved and that deserved intrinsic displeasures are not as bad as otherwise similar displeasures that are not deserved. We observed that on some credible versions of hedonism, atoms that pertain to the intrinsic value of worlds have as elements desert of intrinsic pleasure or of intrinsic displeasure (Ross 1930, Feldman 2004). Many pleasures (or displeasures), if deserved, are deserved on the basis of performing actions. If, in virtue of having worked hard, you deserve some pleasures, the pleasures you deserve are deserved on the basis of your having performed various, pertinent actions. Action-based pleasures (or displeasures) are pleasures (or displeasures) that, if deserved, are deserved on the basis of performing actions.

Such pleasures or displeasures are free only if they causally issue from decisions that are free. As I explained, the rough contours of the argument for this view are these: we have indirect freedom (or control) over something only if we have control over it by virtue of having control over something else. We have direct control over something only if our control over it is not indirect. The control that we or most of us have over our emotion-, feeling-, or attitude-tokens over which we do have control is indirect. On pain of avoiding an infinite regress, all indirect control, eventually, must trace to something over which we have direct control. On my view, we have direct control only over our decisions if we have direct control over anything. If we have direct control solely over our decisions, then the indirect control (or freedom) that we have over our attitudes, such as taking intrinsic pleasure or intrinsic displeasure in things, will, at the end, derive from the direct control (or freedom) that we have over our decisions. To the extent, then, that our decisions are free, to that extent our pertinent attitudes will be free.

It can further be secured, I claimed, that action-based pleasures are deserved only if they are free. In broad strokes, the reasoning here is the following. Imagine that someone heaps praise on you for a decision that causally derives from springs of action with respect to which you are not autonomous, such as surreptitiously implanted desires, beliefs, and values. Presumably, you do not *deserve* the praise. You do not deserve it because it derives from a decision that itself is not free owing to this decision's causally arising from inauthentic actional springs. If, then, one believes that one is not deserving of praise or blame for making a decision that derives from such springs and so from an unfree decision, one should believe, too, that one does not deserve pleasure (or displeasure) on the basis of such an unfree decision. But action-based pleasures or displeasures had on the basis of unfree decisions are not free. Thus, unfree action-based pleasures (or displeasures) are not deserved. It follows that if an action-based pleasure (or displeasure) is deserved, it is free.

Suppose, then, that action-based pleasures or displeasures are free only if they causally derive from decisions that are free, and that they are deserved only if they are free. As such pleasures and displeasures contribute to worldly value, we can show that freedom has an indirect impact on the intrinsic value of worlds by way of having an impact on mental action and desert.

To trace various implications of the freedom presuppositions of judgments concerning worldly value and welfare value for central issues in the free-will debate, I frequently assumed the perspective of the hard incompatibilist. According to hard incompatibilism, provided that we are not agent causes, we lack free will both in worlds that are deterministic and not deterministic. A number of prominent hard incompatibilists may be construed as theorizing that fully appreciating hard incompatibilism reveals that this position is not such a hard one: they have claimed that though hard incompatibilism undermines responsibility, it leaves intact the truth of many other central varieties of normative judgment. I challenged this claim. I argued that hard incompatibilists incur axiological costs: first, lack of free will has a negative impact on personal well-being. Assuming that free pleasures and displeasures, for instance, contribute to personal well-being, there are no such pleasures (or displeasures) if hard incompatibilism is true. Second, lack of free will detrimentally

affects the value of worlds and it does so via two routes. Again, if free pleasures (and displeasures) or free desire satisfactions (and frustrations) contribute to worldly value, there are no such free elements if hard incompatibilism is true. In addition, if (for instance) deserved pleasures, other things being equal, enhance worldly value, and a significant class of pleasures—roughly those pleasures that if deserved are deserved on the basis of performing actions—are deserved only if they are free, then again hard incompatibilism detrimentally affects worldly value owing to hard incompatibilism's undermining desert.

If I am right about freedom's influence on the truth of judgments regarding welfare and worldly value, free-will theorists might well want to examine more closely the influence, if any, of determinism or indeterminism on the truth of other normative appraisals that are essentially connected to, or dependent on, the truth of such axiological judgments.

Appendix A: On Determinism, Randomness, and Desert

In an intriguing, recent paper, Noa Latham argues that determinism and various forms of event-causal libertarian views undermine the truth of important evaluative judgments of desert (Latham, 2004). In this appendix, I summarize and assess this claim.

A.1 Luck and Desert Principles

Latham submits that determinism impugns all (and only) "desert principles." These are principles or claims "entailing that the intrinsic goodness of a person's receiving pleasure or pain depends on the virtue or vice of the person" (2004, p. 154). Latham does not distinguish between sensory pleasures and pains and attitudinal pleasures and displeasures; but nothing in his discussion seems to turn on this distinction. Virtue and vice are taken to be appropriate desert bases for the receipt of pleasure or pain or for the receipt of anything that is a plausible candidate of being something that is intrinsically good or bad—a primary intrinsic good (or evil)—for the person who possesses it (p. 155). Latham proposes that the pleasures and pains referred to in the desert principles may come about unintentionally, for example, by accident, or they may be bestowed intentionally, for instance, by an agent (p. 155). He explains that he is using "virtue," "vice," and its cognates "as the most general evaluative terms that may apply to an action, character trait, agent, life, or segment of a life" (p. 156). Elaborating, he says:

> The judgment of virtue and vice can be based on the goodness and badness, or rightness and wrongness, of the agent's action or actions. And it may take into consideration the agent's motives, her beliefs about the goodness and rightness of her actions, and the extent to which her actions are voluntary or free in various compatibilist senses (p. 156).

If judgments of virtue and vice can be based on, and may take into account, the agent's motives, her beliefs about the goodness and rightness of her actions, and the extent to which her actions are voluntary or free, then it seems that such judgments may well be predicated on considerations of praiseworthiness and blameworthiness. It seems that the agent's being praiseworthy or blameworthy, just as the agent's actions being right or wrong, or good or bad, may be appropriate "desert bases" for

the receipt of pleasure or pain. As examples of desert principles, Latham advances, among others, the following:

DP1: It is intrinsically good that hard work be fittingly rewarded (p. 156).
DP2: It is intrinsically good for vice to meet with a fitting level of pain (p. 156).

The argument for the incompatibility of determinism and desert principles proceeds by way of a thought experiment derived from Plato's Myth of Er. Imagine that lives fully described from birth to death and, thus, fully determined before they are lived, are to be randomly "assigned" to individuals. In some of them, one lives a life of virtue, in others a life of vice. Having lived the assigned life and being judged for it, one then experiences pleasure or pain befitting the virtuousness or viciousness of one's life (p. 159). If one has been virtuous, one receives a fitting amount of pleasure; if one has been vicious, one receives a fitting amount of pain. To assess moral intuitions concerning desert principles which involve judgments of *intrinsic value* and their compatibility with determinism, and to guard against these intuitions being infected by judgments concerning the *instrumental value* of pleasures or pains, Latham stipulates that the pleasures and pains one receives on the basis of being virtuous or vicious be one's final experience and that no one else knows of it so that it has no consequences for anyone (pp. 157–159). Suppose you draw the life of a cunning serial killer. Latham writes:

> This would seem terrifying and unjust in its own right. But now consider how you would react to the further news that after that life to be is over you will endure years of misery in the underworld befitting the viciousness of that life. You are fully aware that many of your evil actions during that life will have a considerable degree of compatibilist freedom, and that there is a sense in which these actions can be said to be caused by you or by your reasons. Yet I think that, given your current helplessness, your natural response, and indeed the natural response of anyone reflecting on the situation, would be to maintain that such torture in the underworld would merely compound the injustice of having to lead that life. I shall call the view that such torture would not be good the basic intuition.

A.1.1 Latham Adds

> Now consider a variant of the thought experiment in which you are told that a complete life is arbitrarily assigned, that it is a very vicious one, and that at some point *during* the life after committing a particularly wicked act the person experiences pain fitting the wickedness of the act, either accidentally or as [an administered punishment]. . . . So the basic intuition generalizes further to intentionally or unintentionally inflicted pains during someone's life, showing one important desert principle to be unacceptable, namely, that *it is intrinsically good for vice to meet with a fitting level of pain*. . . . I think many would hold that it is precisely because in this thought experiment a fully determined life is arbitrarily assigned that they reject this desert principle. The thought experiment thus suggests that determinism would *undermine* this desert principle. . . . And further reflection I think suggests that all desert principles would be undermined by determinism, including the seemingly more plausible distributional principles such as that [DP3] *it is intrinsically better to punish the vicious than the innocent*. For according to the thought experiment, the lives are delivered in a complete package. The physical attributes, such as race, health, and beauty, and the

> psychological, prudential, and moral attributes you are stuck with are all equally arbitrarily assigned. And this suggests that the level of one's virtue and vice is as irrelevant as one's height or beauty to the goodness of pleasures and pains at the end of one's life, and hence to the intrinsic goodness of pleasures and pains incurred during that life (p. 159, note omitted).

Summarizing, a fully determined life is "assigned" to you. In some you live a life of virtue, in others you live a life of vice. If you live a life of virtue, you are appropriately rewarded with pleasure; if you live a life of vice, you are appropriately bestowed with pain. Indeed, we may assume that whatever life you lead, there is a perfect fit between the pleasures and pains (or primary intrinsic goods) that you receive and the pleasures and pains (or primary intrinsic goods) that you would deserve to receive if you deserved to receive anything in that life. But what life you lead is entirely a draw of the cards. Just as the height you attain is a matter of luck and, hence, cannot be a suitable desert base for the receipt of intrinsic goods or evils, so the virtue or vice you exhibit in your assigned life is dictated by luck and cannot be appropriate desert bases for the receipt of pleasure or pain. A life being assigned to you is tantamount to a life that you would live if all events in that life were causally determined. Thus, determinism is incompatible with desert principles such as DP1–DP3.

Latham submits that a similar line of reasoning undermines event-causal libertarian theories that are not wedded to agent-causation. According to such theories, an action is free only if there is some indeterminism at some point or points in its causal trajectory (for elaboration, see Sections 6.8 and 10.3). More generally, Latham widens the discussion to include event-causal libertarian views in this way: he asks "what evaluative implications it would have if the world were fundamentally probabilistic, and thus involved an element of randomness, as described by quantum theory" (p. 161). Latham argues that desert principles would also be undermined if the world were fundamentally probabilistic:

> Imagine that when reading about the life you have been assigned, you discover that it contains various branching points during decision making at which dice are rolled. If at one of these points a six is thrown, your life takes course A, while if less than a six is thrown your life takes course B. It seem implausible that you or anyone reflecting on the experience of pain that follows the life would think *both* that it would be good for you to suffer at the end of your life when it takes course A because a six is thrown, *and* that it would not be good for you to suffer for a life taking course A if it is entirely determined. Nor is it plausible that you or others will think the situation is crucially underdescribed because it has not been specified at precisely which moments during the decision-making process the dice are rolled. Thus desert principles appear to be undermined by probabilistic indeterminism as well as by determinism. At least it is hard to see a plausible rationale for distinguishing the probabilistic from the deterministic cases (p. 161).

A.2 Evaluation of Latham's Position

How might a compatibilist respond to Latham's views? Imagine that the most robust compatibilist conditions (whatever they are) for free action and praiseworthiness and blameworthiness are satisfied in each of the assigned worlds. Caution should

be exercised in not construing such an assignment in a fashion in which compatibilists (and various libertarians) would regard as responsibility-undermining. So, for instance, the assignment is *not* analogous to the "assignment" of desires or beliefs in global induction cases such as the Ann/Beth case. Suppose that in your world of vice, you are morally blameworthy (in the "compatibilist" sense of "blameworthiness") for your evil deeds. You are deserving of blame on the basis of an appropriate desert base, perhaps the base being your intentionally and (freely in the compatibilist sense of "free") having done something that you take to be morally wrong. If you are so *deserving* of blame (in the eyes of the compatibilist), there is little reason why a compatibilist should deny that you may well also deserve fitting amounts of pain on the basis of being blameworthy and that it would be intrinsically good if you did receive fitting amounts of pain. But what of your life's being "dealt" to you randomly and, thus, its being merely a matter of luck what sort of life you live? Well, compatibilists have long acknowledged the phenomenon of what we may call a variety of "remote luck." If determinism is true, there is a sense in which the life you lead is a matter of (remote) luck: distant facts of the past (at a time at which you did not even exist) together with the natural laws "dictate" the sort of life you lead. But compatibilists do not (generally) believe that such remote luck undermines praiseworthiness or blameworthiness. If you can be deserving of blame despite such luck, then there is, again, no reason why a compatibilist should deny that you may, for instance, deserve to receive pain on the basis of intentionally (and freely) doing something that you take to be wrong.

Suppose your assigned world is indeterministic (or probabilistic) and that the most robust event-causal libertarian conditions for free action and praiseworthiness and blameworthiness are satisfied in your world. A libertarian, endorsing the same sort of (suitably adjusted) reasoning outlined in the previous paragraph, should have little reservation in eschewing the claim that, for instance, DP2 is incompatible with indeterminism.

Perhaps Latham's thought experiment is designed, in part, to highlight certain facts about ultimate origination. As we have mentioned, various free will theorists, such as Kane and Pereboom, endorse a principle of this sort:

O: If an agent is morally responsible for deciding to perform an action, then the production of this mental action, a decision, must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control.

"Control," as it occurs in the principle, is ambiguous but the rough idea underlying the principle is pellucid enough. It seems that we do not have relevant control—complete or partial, direct or indirect—over the *distant* past and the laws of nature. If determinism is true, all events, including human actions, are upshots of the distant past and the laws. Hence, if our world is deterministic, we are not the ultimate originators of any of our actions. It would then follow, given *Principle O*, that none of our actions is free and, consequently, we are not morally responsible for any of our actions.

Maybe one of Latham's thoughts is that in an "assigned world," one is not the ultimate originator of one's virtue or vice, or if one wants, one is not the ultimate originator of at least some desert bases (for instance, the time and effort one expends in acquiring a certain good). In a determined world, just as it is false that one is the ultimate originator of one's height in that one has no relevant control over the elements that determine one's height, so it is false that one is the ultimate originator of the factors, whatever they may be, that determine what pertinent capacities one has in that one has no control in or over the acquisition of these factors. If one has the capacity to do good, or the capacity for hard work, one is not the ultimate originator of this capacity. And just as it would be inappropriate to receive pleasure on the basis of one's height owing to one's not being the ultimate originator of height-determining factors, so it would be inappropriate to receive pleasure on the basis of the capacity (or its exercise) to do good or to work hard owing to one's not being the ultimate originator of the capacity-determining factors.

There is a problem, however, with this bit of reasoning as well. A compatibilist may draw attention to the fact that although we may have no relevant control in acquiring various beliefs or desires, nonetheless, once these things are acquired, we may have considerable deterministic control in, for example, assessing some belief, modulating the influence of a belief on deliberation, attenuating the strength of some desire, and so on. Similarly, we may have no relevant control in acquiring factors that are, in some way or another, pertinent to certain desert bases. We may, for instance, have no germane control in acquiring the disposition or capacity to work hard. Still, having acquired the capacity, we may command significant compatibilist control in, for example, exercising this capacity. In addition, there are compatibilist and incompatibilist renditions of ultimate origination. On a compatibilist account of ultimate origination, we may well be the ultimate originators of factors that contribute to relevant desert bases. Thus, the compatibilists need not deny that it would be intrinsically good for a person to receive fitting amounts of pleasure or pain on the basis of these bases even in randomly assigned worlds.

Appendix B: On the Creation of Worlds and Worldly Value

A possible world is supremely best—"S-best"—if and only if its intrinsic value exceeds the intrinsic value of any other world. Leibniz's elegant argument—the "creation argument"—for the view that the actual world is supremely best may be summarized in this way: prior to creating or actualizing (the creatable parts of) any possible world, God's essential omniscience guarantees that God knows the value of each world and so God knows what world is S-best. Existing in each possible world, his essential omnipotence ensures that in each such world, he can create the S-best world. Finally, being essentially omnibenevolent God prefers the S-best world to all other possible worlds. As the actual world is God's creation, it must be that this world is S-best; it is the best of all possible worlds (Leibniz 1951, pp. 267–268). The principle of sufficient reason—on one rendition, the principle that for each contingent thing, there is some independent sufficient reason why it is as it is, rather than otherwise—commits Leibniz to the view that God must have had sufficient reason to create the actual world (see, e.g., Leibniz 1975, pp. 112, 211). I take this reason (or reasons) to be practical. In this appendix, I expose a problem with the creation argument. The problem can be summarized in this way: if freedom affects the value of worlds (as I have argued), and if the sort of freedom at issue is a species of libertarian freedom (as I will explain), then, prior to creation, God cannot know the value of some worlds and, consequently, God cannot know what world is S-best.

Construe "God" as a proper name for an entity who has the divine attributes essentially. Among these attributes are omniscience and eternality. Assume that a being is omniscient only if it has maximal knowledge and that it is eternal in the way in which God (if he exists) is, only if it exists at every time. I assume, further, that there is a supremely best world. This assumption is, of course, controversial. So, for instance, it may be that there is no S-best world but there are only optimals (a world is optimal if and only if there is no world that is intrinsically better than it). Or it may be that for each world, there is a world that is better than it is. But this dispute over whether or not there is a supremely best world is of no consequence to what follows. What *is* of import is that God's choice concerning which world to actualize turns pivotally on God's knowledge of the intrinsic value of that world.

Another assumption that I make is that intrinsic attitudinal pleasures and displeasures affect the intrinsic value of a world. Other things being equal, a world with

greater net intrinsic attitudinal pleasure than another is better than this other world. I have proposed that free pleasures are better than otherwise similar pleasures that are not free and that free displeasures are not as bad as otherwise similar pleasures that are not free. The freedom at issue is responsibility-level freedom—the freedom that responsibility requires. Pleasures (and displeasures), if free, are indirectly free; they derive their freedom from the freedom of pertinent decisions. A libertarian places the following constraint on free decisions. Let *d* stand for Michael's decision to do something—a decision that he makes at some time, *ts*. Let *c* be a proposition that expresses all the causal conditions that obtain in the world up to *ts*. Michael's decision, *d*, that he makes at *ts* is a libertarian free (mental) action only if, possibly, (i) Michael is in *c* and Michael freely refrains, at *ts*, from making *d*, and (ii) if *d* is caused by agent-involving events, such as Michael's having of reasons, *d* is indeterministically caused by these events.[1] In short, if Michael makes a decision at *t* in world *w*, this decision is free in the libertarian sense of "free" at issue only if, there is a possible world with the same laws and the same past up to or just prior to *t* as *w* in which, at *t*, Michael refrains from making this decision. I now make a final assumption: there are free worlds; that is, there are worlds in which agents freely (in the libertarian sense of "freely") take intrinsic pleasure in some things and freely take intrinsic displeasure in other things.

Suppose *w1* is a free world, and suppose that in this world, at *t*,

(c1): Michael freely takes intrinsic attitudinal pleasure in writing a paper,
and
(c2): Michelle freely takes intrinsic attitudinal displeasure in taking some medicine.

How does God, at a time prior to *t*, know propositions, such as (c1) and (c2) that predict the occurrence of undetermined contingent events? Presumably, if God foreknows the future by using a certain faculty or mode of belief acquisition, then that faculty is infallible: necessarily, any belief acquired by this faculty is true. God's use of inductive means to acquire beliefs about future contingent events is, thus, excluded by the requirement of infallibility. William Hasker (1989, p. 56) suggests, plausibly, that God can have access to such events either (a) through infallible deduction from prior nonrelational facts and the natural laws, or (b) via middle knowledge, or (c) by simple foreknowledge: God directly knows actual future contingents, undetermined ones included, apparently through infallible perception.

Each of these options, however, has drawbacks. The Molinist's position (b), whose difficulties have been amply recorded and discussed in the literature, need be considered no further (see, e.g., Adams 1977, Hasker 1989, Hunt 1990, Zagzebski 1991, Hoffman and Rozenkrantz 2002, pp. 120–124). A serious shortcoming of (c) is that simple foreknowledge—that is, infallible foresight of a future

[1] Not all libertarians accept conditions (i) and (ii). Nontraditional libertarians reject the requirement that a decision is free only if one could have refrained from performing it. Libertarians who are traditional agent-causalists endorse the alternative possibilities requirement and insist that a free decision has no event as its cause but has, instead a substance, its agent, as its cause.

contingent event—requires "backward causation." As Joshua Hoffman and Gary Rosenkrantz explain,

> Necessarily, if *S* has infallible foresight of an undetermined future contingent event, *e*, then *S*'s perception of *e* must have as its efficient cause an event which is in the *future* relative to *S*'s perception of *e*, namely, *e*! But it seems to be a necessary truth that if an event e_1 is an efficient cause of an event e_2, then either e_1 is *before* e_2 or e_1 is *simultaneous* with e_2. Thus, if e_1 causes e_2, then this entails that e_1 is not in the future relative to e_2 (2002, p. 119).

This leaves (a) which appears to be the most reasonable of the three options. The occurrence, though, of an *undetermined* future contingent event cannot be deduced from laws and relevant facts. Thus, if (b) and (c) are not in the running or at least suspect, an omniscient being, it seems, cannot have knowledge of undetermined future contingent events.[2]

If, prior to creation, God cannot know (c1), (c2), and relevantly similar propositions, then God cannot know the overall intrinsic value of free worlds such as *w1*. For merely assume that if, in *w1*, at *t*, Michael freely takes pleasure in writing a paper, then *w1* has a certain intrinsic value, *v1*; and that if it is false that in *w1*, at *t*, Michael freely takes pleasure in writing a paper, *w1*'s intrinsic value is something other than *v1*. Since God cannot know propositions that predict the occurrence of future contingent events, prior to creation, God cannot know whether *w1*'s value is *v1* or some other value. It follows that God cannot know the value of free worlds. This, in turn, jeopardizes the creation argument.

[2] For a penetrating defense of this claim, see Hoffman and Rozenkrantz (2002, Chapter 6).

References

Adams, Robert M. 1977. "Middle Knowledge and the Problem of Evil." *American Philosophical Quarterly* 14: 109–117.

Adams, Robert M. 1999. *Finite and Infinite Goods: A Framework for Ethics*. Oxford: Oxford University Press.

Archard, David. 2003. "Children." In *The Oxford Handbook of Practical Ethics.* Hugh LaFollette, ed. Oxford: Oxford University Press.

Aristotle. 1941. *Nicomachean Ethics*. In *The Basic Works of Aristotle*. Richard McKeon, ed. New York: Random House.

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. New York: Oxford University Press.

Audi, Robert. 1982. "Self-Deception, Action, and the Will." *Erkenntnis* 18: 137–138.

Audi, Robert. 1994. "Dispositional Beliefs and Dispositions to Believe." *Nous* 28: 419–434.

Barry, Brian. 1965. *Political Argument*. London: Routledge & Kegan Paul.

Brandt, Richard. B. 1967. "Happiness." In*Encyclopedia of Philosophy.* Paul Edwards, ed. New York: The Macmillan Company & the Free Press (pp. 413–414).

Brandt, Richard. B. 1979. *A Theory of the Good and the Right*. New York: Oxford University Press.

Brandt, Richard. B. 1989. "Fairness to Happiness." *Social Theory & Practice* 15: 33–58.

Brentano, Franz. 1969. *The Origin of Our Knowledge of Right and Wrong*. Translated by Robert Chisholm and Elizabeth Schneewind. London: Routledge & Kegan Paul.

Bykvist, Krister. 2002. "Sumner on Desires and Well-Being." *Canadian Journal of Philosophy* 32: 475–490.

Chisholsm, Roderick. 1986. *Brentano and Intrinsic Value*. Cambridge: Cambridge University Press.

Christman, John. 1991. "Autonomy and Personal History." *Canadian Journal of Philosophy* 21: 1–24.

Clarke, Randolph. 1996. "Agent Causation and Event Causation in the Production of Free Action." *Philosophical Topics* 24: 19–48.

Clarke, Randolph. 2003. *Libertarian Accounts of Free Will.* New York: Oxford University Press.

Clarke, Randolph. 2004. "Reflections on an Argument from Luck." *Philosophical Topics* 32:47–64.

Clarke, Randolph. 2005. "Agent Causation and the Problem of Luck." *Pacific Philosophical Quarterly* 86: 408–421.

Cupit, Geoffrey. 1996. "Desert and Responsibility." *Canadian Journal of Philosophy* 26: 83–100.

Cuypers, Stefaan, E. 2000. "Autonomy Beyond Voluntarism: In Defense of Hierarchy." *Canadian Journal of Philosophy* 30: 225–256.

Davis, W. 1981a. "Pleasure and Happiness." *Philosophical Studies* 39: 305–318.

Davis, W. 1981b. "A Theory of Happiness." *American Philosophical Quarterly* 18: 111–120.

Double, Richard. 2004. "The Ethical Advantages of Free Will Subjectivism." *Philosophy and Phenomenological Research* 69: 411–422.

Feldman, Fred. 1986. *Doing The Best We Can*. Dordrecht: D. Reidel Publishing Company.

Feldman, Fred. 1995. "Desert: Reconsideration of Some Received Wisdom." *Mind* 104: 63–77.

Feldman, Fred. 1996. "Responsibility as a Condition for Desert." *Mind* 105: 165–168.
Feldman, Fred. 1997. *Utilitarianism, Hedonism, and Desert*. Cambridge: Cambridge University Press.
Feldman, Fred. 2000. "Basic Intrinsic Value." *Philosophical Studies* 99: 319–346.
Feldman, Fred. 2004. *Pleasure and the Good Life. Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford: Clarendon Press.
Feldman, Fred. n.d. "What is This Thing Called Happiness?"
Fischer, John Martin and Ravizza, Mark. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.
Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829–839.
Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of A Person," *Journal of Philosophy* 68: 5–20.
Frankfurt, Harry. 1999. "Autonomy, Necessity, and Love." *Necessity, Volition, and Love*. Cambridge: Cambridge University Press (pp. 129–141).
Ginet, Carl. 1990. *On Action*. Cambridge: Cambridge University Press.
Ginet, Carl. 1996. "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing." *Philosophical Perspectives* 10: 403–441.
Ginet, Carl. 1997. "Freedom, Responsibility, and Agency." *Journal of Ethics* 1: 85–98.
Glannon, Walter. 1998. "Moral Responsibility and Personal Identity." *American Philosophical Quarterly* 35: 231–249.
Goetz, Stewart. 1988. "A Noncausal Theory of Agency." *Philosophy and Phenomenological Research* 49: 303–316.
Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.
Haji, Ishtiyaque. 1998. *Moral Appraisability. Puzzles, Proposals, and Perplexities*. New York: Oxford University Press.
Haji, Ishtiyaque. 2000a. "Death and Asymmetries in Moral Appraisals." *Midwest Studies in Philosophy* 24: 135–150.
Haji, Ishtiyaque. 2000b. "Indeterminism, Explanation, and Luck." *Journal of Ethics* 4: 211–235.
Haji, Ishtiyaque. 2002. *Deontic Morality and Control*. Cambridge: Cambridge University Press.
Haji, Ishtiyaque. 2003. "Alternative Possibilities, Luck, and Moral Responsibility." *The Journal of Ethics* 7: 253–275.
Haji, Ishtiyaque. 2004a. "Active Control, Agent-Causation, and Free Action." *Philosophical Explorations* 7: 131–148.
Haji, Ishtiyaque. 2004b. "Freedom, Hedonism, and the Intrinsic Value of Lives." *Philosophical Topics* 32: 131–151.
Haji, Ishtiyaque. 2005a. "Libertarianism, Luck, and Action Explanation." *The Journal of Philosophical Research* 30: 321–340.
Haji, Ishtiyaque. 2005b. "Freedom, Obligation, and Responsibility: Prospects for a Unifying Theory," *Midwest Studies in Philosophy* 29: 106–125.
Haji, Ishtiyaque. 2006. "The Principle of Alternate Possibilities and a Defeated Dilemma." *Philosophical Explorations* 9: 179–202.
Haji, Ishtiyaque. forthcoming. *Incompatibilism's Allure: Principal Arguments for Incompatibilism*. Toronto: Broadview Press.
Haji, Ishtiyaque and Cuypers, Stefaan E. 2004. "Responsibility and the Problem of Manipulation Reconsidered." *International Journal of Philosophical Studies* 12: 439–464.
Haji, Ishtiyaque and Cuypers, Stefaan E. 2006. "Hard- and Soft-Line Responses to Pereboom's Four-Case Manipulation Argument." *Acta Analytica* 21: 19–35.
Haji, Ishtiyaque and Cuypers, Stefaan E. 2007. "Magical Agents, Global Induction, and the Internalism/Externalism Debate." *Australasian Journal of Philosophy* 85: 343–371.
Haji, Ishtiyaque and Cuypers, Stefaan E. 2008. *Moral Responsibility, Authenticity, and Education*. New York: Routledge.
Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford: Clarendon Press.

Harman, Gilbert. 1967. "Toward a Theory of Intrinsic Value." *Journal of Philosophy* 64: 792–804.
Hasker, William. 1989. *God, Time, and Knowledge*. Ithaca: Cornell University Press.
Heathwood, Chris. 2005. "The Problem of Defective Desires." *Australasian Journal of Philosophy* 83: 487–504.
Heathwood, Chris. 2006. "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128: 539–563.
Hoffman, J. and Rozenkrantz, G. S. 2002. *The Divine Attributes*. Oxford: Blackwell.
Honderich, Ted. 1988. *The Consequences of Determinism. A Theory of Determinism, Volume 2*. Oxford: Clarendon Press.
Honderich, Ted. 1993. *How Free Are You? The Determinism Problem.* Oxford: Oxford University Press.
Honderich, Ted. 2002. "Determinism as True, Compatibilism and Incompatibilism as False, and the Real Problem." In *The Oxford Handbook of Free Will.* Robert Kane, ed. New York: Oxford University Press (pp. 461–476).
Hooker, Brad. 1991. "Mark Overvold's Contribution to Philosophy." *Journal of Philosophical Research* 16: 333–444.
Hunt, David. 1990. "Middle Knowledge: The Foreknowledge Defense." *International Journal for Philosophy of Religion* 28: 1–24.
Kagan, Shelly. 1998. *Normative Ethics*. Boulder: Westview Press.
Kahneman, D. 1999. "Objective Happiness." In *Well-Being: The Foundations of Hedonic Psychology*. D. Kahneman, E. Diener and N. Schwarz, eds. New York: Russell Sage Foundation (pp. 3–25).
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.
Kane, Robert. 2000. "Non-Constraining Control and the Threat of Social Conditioning." *The Journal of Ethics* 4: 401–403.
Kekes, John. 1982. "Happiness." *Mind* 91: 358–376.
Kekes, J. 1992. "Happiness." In *Encyclopedia of Ethics.* L. C. Becker and C. B. Becker eds. New York: Garland (pp. 430–435).
Latham, Noa. 2004. "Determinism, Randomness, and Value." *Philosophical Topics* 32: 153–167.
Leibniz, G. W. 1951. *Theodicy*. Edited with an introduction by A. Farrer. Translated by E.M. Huggard. London: Routledge & Kegan Paul.
Leibniz, G. W. 1975. *Philosophical Writings*. Edited by G.H.R. Parkinson. Translated by Mary Morris and G.R. Parkinson. Totowa, N.J.: Rowman & Littlefield.
Lemos, Noah, M. 1994. *Intrinsic Value*. Cambridge: Cambridge University Press.
Lemos, Noah, M. 2006. "Indeterminate Value, Basic Value, and Summation." In *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman*. K. McDaniel, F. R. Raibley, R. Feldman, and M. J. Zimmerman, eds. Aldershot, England: Ashgate Publishing Limited (pp. 71–81).
Lipton, Peter. 1990. "Contrastive Explanation." In *Explanation and Its Limits.* Dudley Knowles, ed. Cambridge: Cambridge University Press (pp. 247–266).
Locke, Don. 1975. "Three Concepts of Free Action I." *Proceedings of the Aristotelian Society*, Sup. Vol. 49: 95–112.
McCann, Hugh. 1998. *The Works of Agency: On Human Action, Will, and Freedom.* Ithaca, N.Y.: Cornell University Press.
McLeod, Owen. 2003. "Desert." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. URL = <http://plato.stanford.edu/archives/fall2003/entries/desert/>.
McLeod, Owen. 2006. "Adjusting Utility for Justice: A Reexamination of the Connections between Desert and Intrinsic Value." In *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman*. K. McDaniel, F. R. Raibley, R. Feldman, and M. J. Zimmerman, eds. Aldershot, England: Ashgate Publishing Limited (pp. 115–133).
Mele, Alfred. 1987. *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control.* New York: Oxford University Press.

Mele, Alfred. 1995. *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.
Mele, Alfred. 1999a. "Ultimate Responsibility and Dumb Luck." *Social Philosophy and Policy* 16: 274–293.
Mele, Alfred. 1999b. "Kane, Luck, and the Significance of Free Will." *Philosophical Explorations* 2: 96–104.
Mele, Alfred. 2003. *Motivation and Agency*. New York: Oxford University Press.
Mele, Alfred. 2004. "Motivated Irrationality." In *The Oxford Handbook of Rationality*. Alfred Mele, ed. New York: Oxford University Press (pp. 240–256).
Mele, Alfred. 2005. "Libertarianism, Luck, and Control." *Pacific Philosophical Quarterly* 86: 381–407.
Mele, Alfred. 2006. *Free Will and Luck*. New York: Oxford University Press.
Mele, Alfred. 2007. "Persisting Intentions." *Nous* 41: 735–757.
Mill, John Stewart. [1861] 1957. *Utilitarianism*. Indianapolis: Bobbs-Merrill.
Miller, David. 1976. *Social Justice*. Oxford: Oxford University Press.
Milo, R. 1984. *Immorality*. Princeton, N.J.: Princeton University Press.
Moore, G. E. [1903] 1962. *Principia Ethica*. Cambridge: Cambridge University Press.
Nagel, Thomas. 1979. *Mortal Questions.* Cambridge: Cambridge University Press.
Noggle, Robert. 1999. "Autonomy and the Paradox of Self-Creation." In *Personal Autonomy. New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. J.S. Taylor, ed. Cambridge: Cambridge University Press (pp. 87–108).
Noggle, Robert. 2002. "Special Agents: Children's Autonomy and Parental Authority." In *The Moral and Political Status of Children*. D. Archard and C. MacLeod, eds. Oxford: Oxford University Press (pp. 97–117).
Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
Nozick, Robert. 1989. *The Examined Life*. New York: Simon and Schuster.
Oakley, Justin. 1992. *Morality and the Emotions*. New York: Routledge.
Olson, Jonas. 2006. "G.E. Moore on Goodness and Reasons." *Australasian Journal of Philosophy* 84: 525–534.
Overvold, Mark Carl. 1980. "Self-Interest and the Concept of Self-Sacrifice." *Canadian Journal of Philosophy* 10: 105–118.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Parfit, Derek. 2001. "Rationality and Reasons." In *Exploring Practical Philosophy: From Action to Values*. D. Egonsson, B. Peterson, J. Josefsson, and T. Ronnow-Rasmussen, eds. Aldershot: Ashgate (pp. 17–39).
Pereboom, Derk. 1995. "Determinism *al Dente*," *Noûs* 29: 21–45.
Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
Pereboom, Derk. 2002. "Living Without Free Will: The Case For Hard Incompatibilism." In *The Oxford Handbook of Free Will*. Robert Kane, ed. New York: Oxford University Press (pp. 477–488).
Rabinowicz, Wlodek, and Osterberg, Jan. 1996. "Value Based on Preferences: On Two Interpretations of Preference Utilitarianism." *Economics and Philosophy* 12: 1–27.
Rabinowicz, Wlodek and Ronnow-Rasmussen, Tonni. 2004. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114: 391–423.
Rabinowicz, Wlodek and Ronnow-Rasmussen, Tonni. 2006. "Buck-Passing and the Right Kind of Reasons." *The Philosophical Quarterly* 56: 114–120.
Rachels, James. 1978. "What People Deserve." In *Justice and Economic Distribution*. John Arthur and William H. Shaw, eds. Englewood Cliffs: Prentice-Hall (pp. 150–163).
Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
Rescher, Nicholas. 1966. *Distributive Justice*. New York: Bobbs-Merrill.
Ronnow-Rasmussen, Tonni. 2002a. "Hedonism, Preferentialism, and Value Bearers." *The Journal of Value Inquiry* 36: 463–472.

Ronnow-Rasmussen, Tonni. 2002b. "Instrumental Values—Strong and Weak." *Ethical Theory and Moral Practice* 5: 23–43.
Ross, W.D. 1930. *The Right and the Good.* Oxford: Oxford University Press.
Sadurski, Wojciech. 1985. *Giving Desert Its Due: Social Justice and Legal Theory*. Dordrecht: D. Reidel Publishing Co.
Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
Sen, Amartya. 1987. *On Ethics and Economics*. Oxford: Basil Blackwell.
Smart, J.J.C. 1973. "An Outline of a System of Utilitarian Ethics." In *Utilitarianism: For and Against*, by J.J.C. Smart and Bernard Williams. Cambridge: Cambridge University Press (pp. 3–74).
Stratton-Lake, P. 2005. "How to Deal with Evil Demons: Comment on Rabinowicz and Ronnow-Rasmussen." *Ethics* 115: 788–798.
Strawson, Peter. F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.
Sumner, L. W. 1996. *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
Sumner, L. W. 2006. "Feldman's Hedonism." In *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman*. K. McDaniel, F. R. Raibley, R. Feldman, and M. J. Zimmerman, eds. Aldershot, England: Ashgate Publishing Limited (pp. 83–100).
Tatarkiewicz, Wladyslaw. 1966. "Happiness and Time." *Philosophy and Phenomenological Research* 27: 1–10.
Tatarkiewicz, Wladyslaw. 1976. *Analysis of Happiness*. The Hague: Martinus Nijhoff.
Telfer, Elizabeth. 1980. *Happiness*. New York: St. Martin's Press.
van Inwagen, Peter. 1983. *An Essay on Free Will.* Oxford: Oxford University Press.
Vayrynen, Pekka. 2006. "Resisting the Buck-Passing Account of Value." In *Oxford Studies in Metaethics: Volume II*. R. Shafer-Landau, ed. New York: Oxford University Press (pp. 295–324).
Waller, Bruce. 1990. *Freedom Without Responsibility*. Philadelphia: Temple University Press.
Zagzebski, Linda. 1991. *The Dilemma of Freedom and Foreknowledge*. New York: Oxford University Press.
Zimmerman, Michael J. 1996. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.
Zimmerman, Michael J.1997. "A Plea for Accuses." *American Philosophical Quarterly* 34: 229–243.
Zimmerman, Michael J. 2001. *The Nature of Intrinsic Value*. Lanham, Maryland: Rowman & Littlefield.
Zimmerman, Michael, J. 2005. "Review of Deontic Morality and Control." *Philosophy and Phenomenological Research* 70: 492–495.
Zimmerman, Michael. J. 2006. "Moral Luck: A Partial Map." *Canadian Journal of Philosophy* 36: 585–608.
Zimmerman, Michael. J. 2007. "Feldman on the Nature and Value of Pleasure." *Philosophical Studies* 136: 425–437.

Index