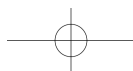
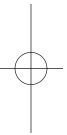
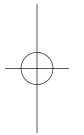


**DRAFT**

# Exploring the Illusion of Free Will and Moral Responsibility



**DRAFT**



**DRAFT**

# Exploring the Illusion of Free Will and Moral Responsibility

Edited by Gregg D. Caruso

LEXINGTON BOOKS  
Lanham • Boulder • New York • Toronto • Plymouth, UK

**DRAFT**

Published by Lexington Books  
A wholly owned subsidiary of The Rowman & Littlefield Publishing Group, Inc.  
4501 Forbes Boulevard, Suite 200, Lanham, Maryland 20706  
www.rowman.com

10 Thornbury Road, Plymouth PL6 7PP, United Kingdom

Copyright © 2013 by Lexington Books

*All rights reserved.* No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the publisher, except by a reviewer who may quote passages in a review.

British Library Cataloguing in Publication Information Available

**Library of Congress Cataloging-in-Publication Data**

TO COME.

Caruso, Gregg D. ed. Exploring the Illusion of Free Will and Moral Responsibility.  
p. cm.

Includes bibliographical references and index.


ISBN 978-0-7391-7731-0 (cloth : alk. paper) -- ISBN 978-0-7391-7732-7 (electronic)

TEST 1. Aesthetics. 2. Postmodernism. I. Rundell, John F. II. Title.

TEST BH39.H445 2011

TEST 111'.85--dc22

TEST 2010037457

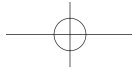
™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48-1992.

Printed in the United States of America

**DRAFT**

# Contents

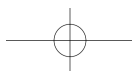
Acknowledgments	vii
Introduction: Exploring the Illusion of Free Will and Moral Responsibility <i>Gregg D. Caruso</i>	1
<b>I: Philosophical Explorations: Free Will Skepticism and Its Implications</b>	<b>17</b>
1 Skepticism about Free Will <i>Derk Pereboom</i>	19
2 The Impossibility of Ultimate Responsibility? <i>Galen Strawson</i>	41
3 Determinism, Incompatibilism and Compatibilism, Actual Consciousness and Subjective Physical Worlds, Humanity <i>Ted Honderich</i>	53
4 The Stubborn Illusion of Moral Responsibility <i>Bruce Waller</i>	65
5 Be a Skeptic, Not a Metaskeptic <i>Neil Levy</i>	87
6 Free Will as a Case of “Crazy Ethics” <i>Saul Smilansky</i>	103
7 The Potential Dark Side of Believing in Free Will (and Related Concepts): Some Preliminary Findings <i>Thomas Nadelhoffer and Daniela Goya Tocchetto</i>	121
8 The People Problem <i>Benjamin Vilhauer</i>	141



*Contents*

**DRAFT**

<b>9</b>	Living Without Free Will <i>Susan Blackmore</i>	161
<b>10</b>	If Free Will Doesn't Exist, Neither Does Water <i>Manuel Vargas</i>	177
<b>11</b>	Free Will and Error <i>Shaun Nichols</i>	203
<b>II: Scientific Explorations: The Behavioral, Cognitive, and Neurosciences</b>		<b>219</b>
<b>12</b>	The Complex Network of Intentions <i>John-Dylan Haynes and Michael Pauen</i>	221
<b>13</b>	Experience and Autonomy: Why Consciousness Does and Doesn't Matter <i>Thomas W. Clark</i>	239
<b>14</b>	What Does the Brain Know and When Does It Know It? <i>Mark Hallett, MD</i>	255
<b>15</b>	If Free Will Did Not Exist, It Would Be Necessary to Invent It <i>Susan Pockett</i>	265
<b>16</b>	Free Will, an Illusion?: An Answer from a Pragmatic Sentimentalist Point of View <i>Maureen Sie</i>	273
	References	291
	About the Contributors	311

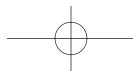
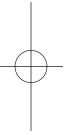
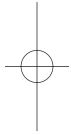


**DRAFT**

## Acknowledgments

**[B02.0]** This volume contains all original essays specially written for this collection with the exception of Galen Strawson’s “The Impossibility of Ultimate Responsibility?” which was first published in *Free Will and Modern Science*, edited by Richard Swinburne (London: British Academy, 2011). It is reprinted here with the permission of the author and the British Academy. I would like to thank all of the contributors for their participation and hard work: Susan Blackmore, Thomas W. Clark, Mark Hallett, John-Dylan Haynes, Ted Honderich, Neil Levy, Thomas Nadelhoffer, Shaun Nichols, Michael Pauen, Derk Pereboom, Susan Pockett, Maureen Sie, Saul Smilansky, Galen Strawson, Daniela Goya Tocchetto, Manuel Vargas, Benjamin Vilhauer, and Bruce Waller. I would also like to thank Corning Community College for granting a sabbatical to work on this project as well as the encouragement I received from the following friends and family members: Liz Vlahos, Robert Talisse, Rob Tempio, David Rosenthal, Michael Levin, Josh Weisberg, Christine Atkins, Andrea Harris, James Cruz, John Celentano, the Hathaway family, Louis and Dolores Caruso, my brothers, brothers-in-law, their families, and George and Marina Kokkinos.

**[B02.1]** Lastly, I would like to thank my wife, Elaini, and my daughter, Maya, for their never ending supply of love and support—I cannot thank you enough!





**DRAFT**

# Introduction

## *Exploring the Illusion of Free Will and Moral Responsibility*

Gregg D. Caruso

**[B03.0]** This book is aimed at readers who wish to explore the philosophical and scientific arguments for free will skepticism and their implications.<sup>1</sup> Skepticism about free will and moral responsibility has been on the rise in recent years. In fact, a significant number of philosophers, psychologists, and neuroscientists now either doubt or outright deny the existence of free will and/or moral responsibility—and the list of prominent skeptics appears to grow by the day.<sup>2</sup> Given the profound importance that the concepts of free will and moral responsibility play in our lives—in understanding ourselves, society, and the law—it is important that we explore what is behind this new wave of skepticism. It is also important that we explore the potential consequences of skepticism for ourselves and society. That is what this volume attempts to do. It brings together an internationally recognized line-up of contributors, most of whom hold skeptical positions of some sort, to display and explore the leading arguments for free will skepticism and to debate their implications.

### **[B03.1]** FREE WILL SKEPTICISM AND ITS IMPLICATIONS

**[B03.2]** Contemporary theories of free will tend to fall into one of two general categories, namely, those that insist on and those that are skeptical about the reality of human freedom and moral responsibility.<sup>3</sup> The former category includes *libertarian* and *compatibilist* accounts of free will, two general views that defend the reality of free will but disagree on its nature. The latter

category includes a family of skeptical views that all take seriously the possibility that human beings do not have free will, and are therefore not morally responsible for their actions in a way that would make them *truly deserving* of blame and praise for them. The main dividing line between the two pro-free will positions, libertarianism and compatibilism, is best understood in terms of the traditional problem of free will and determinism. *Determinism*, as it is commonly understood, is roughly the thesis that every event or action, including human action, is the inevitable result of preceding events and actions and the laws of nature.<sup>4</sup> The problem of free will and determinism therefore comes in trying to reconcile our intuitive sense of free will with the idea that our choices and actions may be causally determined by impersonal forces over which we have no ultimate control.

Libertarians and compatibilists react to this problem in different ways. Libertarians acknowledge that if determinism is true, and all of our actions are causally necessitated by antecedent circumstances, we lack free will and moral responsibility. Yet they further maintain that at least some of our choices and actions must be free in the sense that they are not causally determined. Libertarians therefore reject determinism and defend a counter-causal conception of free will in order to save what they believe are necessary conditions for free will—i.e., the *ability to do otherwise* in exactly the same set of conditions and the idea that we remain, in some important sense, the *ultimate source/originator* of action. Compatibilists, on the other hand, set out to defend a less ambitious form of free will, one which can be reconciled with the acceptance of determinism. They hold that what is of utmost importance is not the falsity of determinism, nor that our actions are uncaused, but that our actions are voluntary, free from constraint and compulsion, and caused in the appropriate way. Different compatibilist accounts spell out the exact requirements for compatibilist freedom differently but popular theories tend to focus on such things as reasons-responsiveness, guidance control, hierarchical integration, and approval of one's motivational states.<sup>5</sup>

In contrast to these pro-free will positions are those views that either doubt or outright deny the existence of free will and/or moral responsibility. Such views are often referred to as skeptical views, or simply *free will skepticism*, and are the focus of this collection. In the past, the standard argument for skepticism was *hard determinism*: the view that determinism is true, and incompatible with free will and moral responsibility—either because it precludes the *ability to do otherwise* (leeway incompatibilism) or because it is inconsistent with one's being the “ultimate source” of action (source incompatibilism)—hence, no free will. For hard determinists, libertarian free will is an impossibility because human actions are part of a fully deterministic world and compatibilism is operating in bad faith.



**DRAFT**

*Introduction*

**[B03.5]**

Hard determinism had its classic statement in the time when Newtonian physics reigned (see, e.g., d’Holbach 1770), but it has very few defenders today—largely because the standard interpretation of quantum mechanics has been taken by many to undermine, or at least throw into doubt, the thesis of universal determinism. This is not to say that determinism has been refuted or falsified by modern physics, because it has not. Determinism still has its modern defenders, most notably Ted Honderich (1988, 2002), and the final interpretation of physics is not yet in. It is also important to keep in mind that even if we allow some indeterminacy to exist at the microlevel of our existence—the level studied by quantum mechanics—there would still likely remain *determinism-where-it-matters* (Honderich 2002, 5). As Honderich argues: “At the ordinary level of choices and actions, and even ordinary electrochemical activity in our brains, causal laws govern what happens. It’s all cause and effect in what you might call real life” (2002, 5). Nonetheless, most contemporary skeptics defend positions that are best seen as successors to traditional hard determinism.

**[B03.6]**

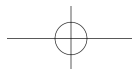
In recent years, for example, several contemporary philosophers have offered arguments for free will skepticism, and/or skepticism about moral responsibility, that are agnostic about determinism—e.g., Derk Pereboom (2001), Galen Strawson (1986/2010), Saul Smilansky (2000), Neil Levy (2011), Richard Double (1991), Bruce Waller (2011), and Gregg Caruso (2012).<sup>6</sup> Most maintain that while determinism is incompatible with free will and moral responsibility, so too is *indeterminism*, especially the variety posited by quantum mechanics. Others argue that regardless of the causal structure of the universe, we lack free will and moral responsibility because free will is incompatible with the pervasiveness of *luck* (Levy 2011). Others (still) argue that free will and ultimate moral responsibility are incoherent concepts, since to be free in the sense required for ultimate moral responsibility we would have to be *causa sui* (or “cause of oneself”) and this is impossible (see Strawson ch.2). Here, for example, is Nietzsche on the *causa sui*:

**[B03.7]**

The *causa sui* is the best self-contradiction that has been conceived so far; it is a sort of rape and perversion of logic. But the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. The desire for “freedom of the will” in the superlative metaphysical sense, which still holds sway, unfortunately, in the minds of the half-educated; the desire to bear the entire and ultimate responsibility for one’s actions oneself, and to absolve God, the world, ancestors, chance, and society involves nothing less than to be precisely this *causa sui* and, with more than Baron Munchhausen’s audacity, to pull oneself up into existence by the hair, out of the swamps of nothingness. (1992, 218-19)<sup>7</sup>

**[B03.8]**

What all these skeptical arguments have in common, and what they share with classical hard determinism, is the belief that what we do, and the way



we are, is ultimately the result of factors beyond our control and because of this we are never morally responsible for our actions in the basic desert sense—the sense that would make us *truly deserving* of blame or praise.<sup>8</sup> This is not to say that there are not other conceptions of responsibility that can be reconciled with determinism, chance, or luck. Nor is it to deny that there may be good pragmatic reasons to maintain certain systems of punishment and reward. Rather, it is to insist that to hold people *truly* or *ultimately* morally responsible for their actions—i.e., to hold them responsible in a non-consequentialist desert-based sense—would be to hold them responsible for the results of the morally arbitrary, for what is ultimately beyond their control, which is (according to these views) fundamentally unfair and unjust.

In addition to these philosophical arguments, there have also been recent developments in the behavioral, cognitive, and neurosciences that have caused many to take free will skepticism seriously. Chief among them have been the neuroscientific discovery that unconscious brain activity causally initiates action prior to the conscious awareness of the intention to act (e.g., Benjamin Libet, John-Dylan Haynes), Daniel Wegner’s work on the double dissociation of the experience of conscious will, and recent findings in psychology and social psychology on *automaticity*, *situationism*, and the *adaptive unconscious* (e.g., John Bargh, Timothy Wilson).<sup>9</sup> Viewed collectively, these developments indicate that much of what we do takes place at an automatic and unaware level and that our commonsense belief that we consciously initiate and control action may be mistaken. They also indicate that the causes that move us are often less transparent to ourselves than we might assume—diverging in many cases from the conscious reasons we provide to explain and/or justify our actions. These findings reveal that the higher mental processes that have traditionally served as quintessential examples of “free will”—such as goal pursuits, evaluation and judgment, reasoning and problem solving, interpersonal behavior, and action initiation and control—can and often do occur in the absence of conscious choice or guidance (Bargh and Ferguson 2000, 926). They also reveal just how wide open our internal psychological processes are to the influence of external stimuli and events in our immediate environment, without knowledge or awareness of such influence. For many these findings represent a serious threat to our everyday folk understanding of ourselves as conscious, rational, responsible agents, since they indicate that the conscious mind exercises less control over our behavior than we have traditionally assumed.

Even some compatibilists now admit that because of these behavioral, cognitive, and neuroscientific findings “free will is at best an occasional phenomenon” (Baumeister 2008b, 17; see also Nahmias forthcoming-a). This is an important concession because it acknowledges that the *threat of shrinking agency*—as Thomas Nadelhoffer (2011) calls it—remains a serious one independent of any traditional concerns over determinism. That is, *even*

**DRAFT**

*Introduction*

if one believes free will and causal determinism can be reconciled, the deflationary view of consciousness which emerges from these empirical findings must still be confronted, including the fact that we often lack transparent awareness of our true motivational states. Such a deflationary view of consciousness is potentially agency undermining (see, e.g., Nadelhoffer 2011; King and Carruthers 2012; Caruso 2012; Sie and Wouters 2010; and Davies 2009) and must be dealt with independent of, and in addition to, the traditional compatibilist/incompatibilist debate.

[B03.11] In addition to these specific concerns over conscious volition and the threat of shrinking agency there is also the more general insight, more threatening to libertarianism than compatibilism, that as the brain sciences progress and we better understand the mechanisms that undergird human behavior, the more it becomes obvious that we lack what Tom Clark (ch.13) calls “soul control.” There is no longer any reason to believe in a non-physical self which controls action and is liberated from the deterministic laws of nature; a little *uncaused causer* capable of exercising counter-causal free will. While most naturalistically inclined philosophers, including most compatibilists, have long given up on the idea of soul control, eliminating such thinking from our folk psychological attitudes may not be so easy and may come at a cost for some. There is some evidence, for example, that we are “natural born” dualists (Bloom 2004) and that, at least in the United States, a majority of adults continue to believe in a non-physical soul that governs behavior (Nadelhoffer in press). To whatever extent, then, such dualistic thinking is present in our folk psychological attitudes about free will and moral responsibility, it is likely to come under pressure and require some revision as the brain sciences advance and this information reaches the general public.<sup>10</sup>

[B03.12] What, then, would be the consequence of accepting free will skepticism? What if we came to disbelieve in free will and moral responsibility? What would this mean for ourselves? Our interpersonal relationships? Society? Morality? The law? What would it do to our standing as human beings? Would it cause nihilism and despair as some maintain? Or perhaps increase anti-social behavior as some recent studies have suggested (Vohs and Schooler 2008; Baumeister, Masicampo, and DeWall 2009)? Or would it rather have a humanizing effect on our practices and policies, freeing us from the negative effects of free will belief? These questions are of profound pragmatic importance and should be of interest independent of the metaphysical debate over free will. As public proclamations of skepticism continue to rise, and as the mass media continues to run headlines announcing “Free will is an illusion” and “Scientists say free will probably doesn’t exist...”<sup>11</sup> we need to ask what effects this will have on the general public and what the responsibility is of professionals.

In recent years a small industry has actually grown up around precisely these questions. In the skeptical community, for example, a number of different positions have been developed and advanced—including Saul Smilansky’s *illusionism* (2000), Thomas Nadelhoffer’s *disillusionism* (2011), Shaun Nichols’ *anti-revolution* (2007), and the *optimistic skepticism* of Derk Pereboom (2001, 2013a), Bruce Waller (2011), Tamler Sommers (2005, 2007b), and others.

[B03.13]

Saul Smilansky, for example, maintains that our commonplace beliefs in libertarian free will and desert-entailing ultimate moral responsibility are illusions,<sup>12</sup> but he also maintains that if people were to accept this truth there would be wide-reaching negative intrapersonal and interpersonal consequences. According to Smilansky, “Most people not only believe in actual possibilities and the ability to transcend circumstances, but have distinct and strong beliefs that libertarian free will is a condition for moral responsibility, which is in turn a condition for just reward and punishment” (2000, 26-27). It would be devastating, he warns, if we were to destroy such beliefs: “the difficulties caused by the absence of ultimate-level grounding are likely to be great, generating acute psychological discomfort for many people and threatening morality—if, that is, we do not have illusion at our disposal” (2000, 166). To avoid any deleterious social and personal consequences, then, and to prevent the unraveling of our moral fabric, Smilansky recommends *free will illusionism*. According to illusionism, people should be allowed their positive illusion of libertarian free will and with it ultimate moral responsibility; we should not take these away from people, and those of us who have already been disenchanted ought to simply keep the truth to ourselves (see also ch.6).

[B03.14]

In direct contrast to Smilansky’s illusionism, Thomas Nadelhoffer defends *free will disillusionism*: “the view that to the extent that folk intuitions and beliefs about the nature of human cognition and moral responsibility are mistaken, philosophers and psychologists ought to do their part to educate the public—especially when their mistaken beliefs arguably fuel a number of unhealthy emotions and attitudes such as revenge, hatred, intolerance, lack of empathy, etc.” (2011, 184). According to Nadelhoffer, “humanity must get beyond this maladaptive suit of emotions if we are to survive.” And he adds, “To the extent that future developments in the sciences of the mind can bring us one step closer to that goal—by giving us a newfound appreciation for the limits of human cognition and agency—I welcome them with open arms” (2011, 184).

[B03.15]

A policy of disillusionism is also present in the optimistic skepticisms of Derk Pereboom, Bruce Waller, Tamler Sommers, and Susan Blackmore. Derk Pereboom, for example, has defended the view that morality, meaning, and value remain intact even if we are not morally responsible, and furthermore, that adopting this perspective could provide significant benefits for our

[B03.16]

lives. In *Living Without Free Will* (2001), he argues that life without free will and desert-based moral responsibility would not be as destructive as many people believe. Prospects of finding meaning in life or of sustaining good interpersonal relationships, for example, would not be threatened (2001, ch.7). And although retributivism and severe punishment, such as the death penalty, would be ruled out, preventive detention and rehabilitation programs would be justified (2001, ch.6). He even argues that relinquishing our belief in free will might well improve our well-being and our relationships to others since it would tend to eradicate an often destructive form of “moral anger.”

[B03.17] Bruce Waller has also made a strong case for the benefits of a world without moral responsibility. In his recent book, *Against Moral Responsibility* (2011), he cites many instances in which moral responsibility practices are counterproductive from a practical and humanitarian standpoint—notably in how they stifle personal development, encourage punitive excess in criminal justice, and perpetuate social and economic inequalities (see Clark 2012 review). Waller suggests that if we abandon moral responsibility “we can look more clearly at the causes and more deeply into the systems that shape individuals and their behavior” (2011, 287), and this will allow us to adopt more humane and effective interpersonal attitudes and approaches to education, criminal justice, and social policy. He maintains that in the absence of moral responsibility, “it is possible to look more deeply at the influences of social systems and situations” (2011, 286), to minimize the patent unfairness that luck deals out in life, and to “move beyond [the harmful effects of] blame and shame” (2011, 287).<sup>13</sup>

[B03.18] In contrast to all these views are those philosophers who argue that as a purely descriptive matter our lives would remain relatively unchanged, not getting better *or* worse, if we were to accept a skeptical or hard determinist perspective. Shaun Nichols, for example, has argued that “people will pretty much stick with the status quo” (2007, 406) when it comes to their everyday interactions and the *reactive attitudes* (P.F. Strawson 1962). He writes, for example: “If people come to accept determinism, what will happen? Opinions on this question differ radically. Some maintain that this would usher in a badly needed revolution in our practices. Others worry that the recognition of determinism would lead to catastrophe. I have a more humdrum guess—if people come to accept determinism, things will remain pretty much the same” (2007, 406). While this is not a defense of free will or compatibilism—for Nichols assumes incompatibilism is intuitive (2007, 405)—it is an interesting descriptive thesis, one which is at odds with the above positions.<sup>14</sup>

[B03.19] Who then is correct? What would the actual consequences of embracing skepticism be? What *should* they be? *Should* we reconsider our attitudes and policies in light of the philosophical and scientific arguments for free will skepticism? These remain important questions, as does the central question: Is skepticism about free will and moral responsibility justified on either

philosophical and/or scientific grounds? Hopefully this collection will aid readers in thinking through these questions, as well as a variety of other issues surrounding free will skepticism and its implications.

## CONTRIBUTIONS

[B03.20]

The sixteen chapters to follow are divided into two main parts. Part I explores the philosophical arguments for free will skepticism and their implications (along with some related issues), while Part II explores recent developments in the behavioral, cognitive, and neurosciences and what they mean for human agency, free will, and moral responsibility.

[B03.21]

In chapter 1, Derk Pereboom presents his argument for free will skepticism known as *hard incompatibilism*. Against the view that free will is compatible with the causal determination of our actions by natural factors beyond our control, he argues that there is no relevant difference between this prospect and our actions being causally determined by manipulators. Against event causal libertarianism, he advances the disappearing agent objection, according to which on this view the agent cannot settle whether a decision occurs, and hence cannot have the control required for moral responsibility. According to Pereboom, a non-causal libertarianism has no plausible proposal of the control in action required for responsibility. He goes on to argue that while agent causal libertarianism may supply this sort of control, it cannot be reconciled with our best physical theories. Since this exhausts the options for views on which we have the sort of free will at issue, he concludes that free will skepticism is the only remaining position. Finally, Pereboom defends the optimistic view that conceiving of life without free will would not be devastating to our conceptions of agency, morality, and meaning in life, and in certain respects it may even be beneficial.

[B03.22]

In chapter 2, Galen Strawson recasts his now famous *Basic Argument* against the possibility of ultimate moral responsibility. He argues that we are not truly or ultimately morally responsible for our actions because we are not *causa sui*. The central idea behind the Basic Argument is that: (1) Nothing can be *causa sui*—nothing can be the cause of itself; (2) In order to be truly or ultimately morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects; (3) Therefore no one can be truly or ultimately morally responsible. Strawson presents several restatements of the Basic Argument along the way, fleshing it out in more detail. He also argues that the Basic Argument cannot simply be dismissed because the idea that we are *causa sui* is a common one and is central to our understanding of ultimate moral responsibility. He concludes by considering possible compatibilist and libertarian responses to the Basic Argument.

[B03.23]



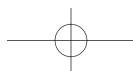
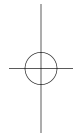
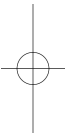


**DRAFT**

*Introduction*

**[B03.24]** In chapter 3, Ted Honderich defends the theory of determinism and explores its consequences. He argues that if we are good empiricists we should accept determinism as true or at least probable—“there ought not be any [reflective empiricists] who are not also determinists.” He also argues that quantum mechanics has not falsified determinism. In fact, the standard interpretation of quantum mechanics, he argues, is a “logical mess” and contains “contradiction” in it. He goes on to explore the consequences of determinism for our lives and for free will. He argues that both compatibilist and incompatibilist approaches fail to adequately deal with the problem of determinism because they both share the mistaken assumption that there is only one conception of free will. Honderich instead argues that there are actually two conceptions of free will—free will as *voluntariness* and as *origination*. While the former is compatible with determinism, the latter is not. Honderich acknowledges, however, that the truth of determinism and the loss of origination create concerns for our “standing” as human beings. In an attempt to preserve some of what is lost when we give up the idea of origination and the responsibility attached to it, Honderich introduces his theory of *actual consciousness*. After briefly sketching how actual consciousness could leave us content despite the loss attendant on giving up origination, Honderich concludes with his “grand hope” for humanity which involves abandoning the “politics of desert” and embracing the Principle of Humanity.

**[B03.25]** In chapter 4, Bruce Waller explores “The Stubborn Illusion of Moral Responsibility.” He begins by pointing out that there is a strange disconnect between the strength of philosophical arguments in support of moral responsibility and the strength of philosophical belief in moral responsibility. While the many arguments in favor of moral responsibility are inventive, subtle, and fascinating, Waller points out that even the most ardent supporters of moral responsibility acknowledge that the arguments in its favor are far from conclusive; and some of the least confident concerning the arguments for moral responsibility—such as Van Inwagen—are most confident of the truth of moral responsibility. Thus, argues Waller, whatever the verdict on the strength of philosophical arguments for moral responsibility, it is clear that belief in moral responsibility—whether among philosophers or the folk—is based on something other than philosophical reasons. He goes on to argue that there are several sources for the strong belief in moral responsibility, but three are particularly influential. First, moral responsibility is based in a powerful “strike back” emotion that we share with other animals. Second, there is a pervasive moral responsibility *system*—extending over criminal justice as well as “common sense”—that makes the truth of moral responsibility seem obvious, and makes challenges to moral responsibility seem incoherent. Finally, there is a deep-rooted “belief in a just world”—a belief that, according to Waller, most philosophers reject when they consciously consider it, but which has a deep nonconscious influence on what we regard



as just treatment and which provides subtle (but mistaken) support for belief in moral responsibility.

In chapter 5, Neil Levy argues that we should be skeptics, not metaskeptics about moral responsibility. In the past, Levy has argued that agents are never morally responsible for their actions because free will and moral responsibility are undermined by luck (2011). In this chapter, he defends his skepticism from a recent challenge presented by Tamler Sommers' metaskepticism. Sommers (2012) denies that there are necessary and sufficient conditions of moral responsibility; instead, he claims, the conditions of moral responsibility vary from culture to culture. Levy admits that if this claim is correct, then no first-order view of moral responsibility, including responsibility skepticism, is true. But, he argues, there are moral reasons to favor skepticism over metaskepticism. Further, he argues that metaskepticism threatens either to collapse into normative nihilism or into first-order skepticism.

**[B03.26]**

In chapter 6, Saul Smilansky introduces the term "crazy ethics" (or CE), which he uses in a semi-descriptive and non-pejorative way to refer to some views that we ourselves hold, or that we think might be true. He claims that some true ethical views are, in this interesting sense, crazy. An ethical view might be considered crazy if it clearly could not serve as a basis for social life. A society that tried to function in the light of such a view would quickly fail. Similarly, an ethical view might be considered crazy if it is self-defeating, so that attempting to implement it would make things worse, in terms of that very view. A view that would be considered far too demanding and thus would be overwhelmingly rejected would also earn the label "crazy"; as would a view that flies in the teeth of fundamental reflective moral beliefs. And an ethical view that needs to be kept apart from the vast majority of the people to whom it applies, an "esoteric morality," also involves craziness. After explicating what makes such views crazy, Smilansky explores the free will problem and attempts to show why viewing it as a case of CE is fruitful. He argues that many of the prevailing positions in the debate are "crazy" in this sense, as are the views he himself holds to be most plausible. He concludes by reflecting on what this means, particularly for morality, personal and social integrity, and the role of philosophy.

**[B03.27]**

In chapter 7, Thomas Nadelhoffer and Daniela Goya Tocchetto explore the potential dark side of believing in free will. They survey some recent findings from moral and political psychology on the possible dark side of believing in free will (and related concepts), as well as report on two exploratory studies of their own, in order to shed some empirical light on the illusionism debate—i.e., the debate over whether we should counsel illusionism (e.g., Smilansky) or disillusionism (e.g., Nadelhoffer). They begin by briefly discussing two of the most recently developed psychometric tools for measuring people's agentic beliefs—namely, the Free Will and Determinism

**[B03.28]**

Scale (FAD-Plus) (Paulhus and Carey 2011) and the Free Will Inventory (FWI) (Nadelhoffer et al. in prep.)—and they explore some of the interesting (and sometimes surprising) correlations that have been found between people’s free will beliefs and their other moral, religious, and political beliefs. They then present and discuss the results of two exploratory studies they conducted to further explore the moral and political psychology of believing in free will. Finally, they attempt to lay the groundwork for future research. They conclude that there is a lot of psychological spadework and philosophical analysis that remains to be done before we can better understand the psychological and social consequences of widespread skepticism about free will.

**[B03.29]** In chapter 8, Benjamin Vilhauer tackles the “People Problem.” He points out that one reason many philosophers are reluctant to seriously contemplate the possibility that we lack free will seems to be the view that we must believe we have free will if we are to regard each other as *persons* in the morally deep sense—the sense that involves deontological notions such as human rights. In the literature of the past few decades, this view is often informed by P.F. Strawson’s view that to treat human beings as having free will is to respond to them with the reactive attitudes, and that if we suspend the reactive attitudes we can only regard human beings as objects to be manipulated in the service of social goals. This purported implication of suspending the reactive attitudes has persuaded many philosophers that we cannot truly treat human beings as persons without assuming that they have free will. Vilhauer argues that this line of thinking is misguided. He agrees with Strawson’s worry that the consequentialism he sees as implicit in the objective attitude could undermine our ability to treat each other as persons, but he argues that it is a mistake to think that we must maintain the reactive attitudes, or any other attributions of free will or moral responsibility, to avoid a depersonalizing slide into consequentialism. He maintains that Kant’s idea of treating people as autonomous ends in themselves, rather than as mere means to ends, provides a compelling analysis of what it means to treat humans as persons, and that there are ways of interpreting Kant’s idea which do not involve reactive attitudes or any other attributions of free will or moral responsibility.

**[B03.30]** In chapter 9, Susan Blackmore defends optimistic skepticism and explores what it is like “Living Without Free Will.” She explains how she came to view free will as an illusion and how she has learned to live without it. She expresses amazement at the many scientists and philosophers who do not believe there is free will but say they still have to live “as if” there is—including several she interviewed for her book *Conversations on Consciousness* (2005). Among the reasons for this contradictory attitude, she speculates, may be fear of letting go of control, fear of losing moral responsibility or acting badly, and fear that society would be impossible or law and order

would break down if we all gave up the illusion. While she can appreciate these fears, she does not share them. She credits science and her thirty years of Zen practice with liberating her from the delusion of free will, and she concludes that living without free will is not only possible but preferable to sticking with the delusion.

In chapter 10, Manuel Vargas questions whether free will skepticism of the varieties usually embraced by scientists are hasty. He begins by arguing that there are three difficulties for such accounts. First, despite frequent appeals to determinism in the work of scientists, it is unclear that determinism is more than a theoretical aspiration in many scientific fields. Second, scientific skeptics too quickly dismiss compatibilist approaches as definitional gambits, rather than as serious accounts that have to be addressed before skepticism carries the day. Third, the powers that are at stake may be high-level, multiply realizable phenomena that resist reduction to the properties that figure in many forms of scientific skepticism. Yet while disputing many of the usual grounds for free will skepticism, Vargas also allows that there are a variety of conceptual and empirical reasons for doubt that we have the powers we ordinarily suppose. Nevertheless, he goes on to articulate an account of how such doubts do not entail the nonexistence of free will. It remains open, he argues, to adopt a revisionist position, where free will is not precisely what ordinary people (or free will skeptics) think it is. On such a view, scientific discoveries about the shape of our agency are not typically reasons for rejecting free will, but rather the basis for better understanding the diverse forms of our freedom and agency.

[B03.31]

In the final chapter of Part I, Shaun Nichols explores the connection between free will and error. He maintains that like other eliminativist arguments in philosophy, arguments that free will is an illusion seem to depend on substantive assumptions about reference. According to free will eliminativists, people have deeply mistaken beliefs about free will and this entails that free will doesn't exist (e.g., Pereboom, Strawson). However, an alternative reaction is that free will does exist, we just have some deeply mistaken beliefs about it (e.g., Vargas). Here Nichols adopts the view that reference is systematically ambiguous. In some contexts, he argues, it is appropriate to take a restrictive view about whether a term embedded in a false theory refers; in other contexts, it's appropriate to take a liberal view about whether a token of the very same term refers. This affords the possibility of saying that the sentence "free will exists" is false in some contexts and true in others. This in turn affords a flexibility in whether we embrace the eliminativist claim. He concludes by arguing that in the case of free will there are practical considerations for and against eliminativism, and that the right conclusion might be a *discretionary* (in)compatibilism.

[B03.32]

In Part II attention shifts to recent developments in the behavioral, cognitive, and neurosciences. It begins, in chapter 12, with neuroscientist John-

[B03.33]

Dylan Haynes and philosopher Michael Pauen questioning whether intentionality is just an illusion. Intentionality is among those features that make humans special. According to many philosophers, intentions are of specific importance when it comes to the difference between mere behavior (stumbling, coughing) and goal-directed action (writing, problem-solving). Recent psychological and neuroscientific evidence, however, has been interpreted as putting the existence of intentions into question. Haynes and Pauen review this evidence and discuss the consequences that it has for our understanding of intentions. They conclude that intentions might well play a role in human action and decision making but that this role differs significantly from what commonsense as well as standard philosophical and folk psychological accounts of intentionality assume.

[B03.34] In chapter 13, Thomas Clark further explores the psychological and neuroscientific evidence and its relationship to consciousness, experience, and autonomy. He maintains that human freedom, responsibility, and autonomy have traditionally been linked to or even identified with conscious control of behavior, where consciousness is widely understood as possibly non-identical with its neural correlates. But the rise of neuroscience strongly suggests that brain processes alone are sufficient for behavior control, and indeed nothing non-physical plays a role in scientific explanations of behavior. Clark addresses three worries that arise in response to investigations of consciousness: the philosophical worry about mental causation; the practical worry about the influence of unconscious processes; and the existential worry that, absent a contra-causal conscious controller, we lack freedom, responsibility, and autonomy. These worries can be defused, Clark argues, by: (1) acknowledging the causal powers of brain-based conscious capacities, those associated with (but perhaps not identical to) conscious experience; (2) expanding the reach of conscious capacities by understanding their limitations; and (3) naturalizing our conceptions of freedom and autonomy. On Clark's account, we need to move beyond the idea of "soul control" to a suitably naturalized understanding of autonomy, one that allows us to hold agents responsible—"at least in the consequentialist sense of being answerable to a moral community."

[B03.35] In chapter 14, neurologist and clinical neurophysiologist Mark Hallett asks "What does the brain know and when does it know it?" He begins his chapter with a brief review of what free will means, and he tries to eliminate any distinction between mind and brain which continues to muddle thinking about the subject. He then turns his attention toward the *qualia* that compose the sense of free will, that of *volition* itself and that of *agency*. The physiology of the qualia of volition and agency are then discussed with a focus on their timing. Hallett explains the neuroscientific findings of Libet et al. (1983), Soon et al. (2008), Lau et al. (2007), as well as findings of his own, and he questions whether volition can be a factor in movement decision or

initiation. He concludes that “by timing arguments alone, it does not appear that the consciousness of willing has any influence, since it comes relatively late in the movement generation process.” Hallett concludes his chapter by considering the implications of this physiology for responsibility.

In chapter 15, Susan Pockett argues that “If free will did not exist, it would be necessary to invent it.” She launches her essay from the cosmologically narcissistic position that it would be a bad thing if the human race destroyed itself, an outcome which is increasingly possible in a technological sense. She argues that ideas matter, and that the concepts of free will and moral responsibility are the most useful tools we have for prevention of such a disaster. Thus, although it may be perfectly acceptable in an intellectual sense to conclude that free will and moral responsibility are merely illusions, teleologically such a conclusion is undesirable to the point of being dangerous. She then goes on to examine two major lines of evidence for the illusory nature of free will. The first line is that neuroscientific experiments show voluntary acts to be neither initiated nor controlled by consciousness. After a brief methodological discussion of the original and still most widely cited experiment of this group, she accepts the conclusion that actions are initiated unconsciously, but argues that free will does not have to involve the conscious initiation of actions. The second major argument for the illusory nature of at least incompatibilist free will is that the success of Western science proves the truth of determinism. She argues that this is not the case: determinism is not a proven fact and very possibly never can be. She therefore concludes that a version of incompatibilist free will which does not require conscious initiation of actions is not necessarily an illusion, and that reports of its death are greatly exaggerated.

**[B03.36]**

In the final chapter, Maureen Sie investigates whether findings in the behavioral, cognitive, and neurosciences are relevant to the concept of free will when that concept is approached, not from a “metaphysical” perspective, but from a different angle—what she labels the *pragmatic sentimentalist* approach (PS-approach). Contrary to the metaphysical approach, the PS-approach does not understand free will as a concept that somehow precedes our moral practices. Rather it is assumed that everyday talk of free will naturally arises in a practice that is characterized by certain reactive attitudes that we take towards one another. First, she explains the social function of moral responsibility that is at the core of the PS-approach. Secondly, she explains how the exchange of reasons central to that social function give rise to a so-called space of reasons. Finally, she examines the scientific findings of recent decades, especially those in psychology and social psychology, that are most relevant to free will understood from the PS-approach. She concentrates on those findings that show that we (1) lack agential transparency, i.e., immediate and infallible introspective access to the motivational origin of our actions and that, as a result, (2) we are sometimes “mistaken” in our

**[B03.37]**



**DRAFT**

*Introduction*

understanding of our own actions. She concludes that “this body of research should lead to serious reconsideration of the occasions on which we claim, e.g., ‘to act on the basis of reasons and out of our own free will . . .’” She also concludes, however, that given the social function of moral responsibility ascriptions and the role of the concept of free will, the claim that it is an illusion makes no sense from the PS-approach.

**[B03.38]**

NOTES

**[B03n1]**

1. I am very grateful to Bruce Waller and Benjamin Vilhauer for helpful comments on earlier drafts of this introduction.

**[B03n2]**

2. For an impressive list of scientists who have recently proclaimed skeptical positions, see Snyder (2012). Also helpful in tracking the rise of free will skepticism is the website run by George Ortega: [www.causalconsciousness.com](http://www.causalconsciousness.com).

**[B03n3]**

3. Most contemporary philosophers argue that *free will* and *moral responsibility* stand or fall together. Exceptions include John Martin Fisher (1994) and Bruce Waller (2011), but such views remain controversial. In fact, much of the philosophical tradition has simply defined “free will” as “a kind of power or ability to make decisions of the sort for which one can be morally responsible” (Fisher, Kane, Pereboom, and Vargas 2007, 1).

**[B03n4]**

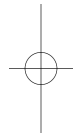
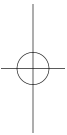
4. Traditional scientific determinism maintains that the state of the universe at any given time is wholly and unequivocally determined by the state of the universe at prior times and the laws of nature. Such determinism is sometimes illustrated by the thought experiment of *Laplace’s demon*—an all-knowing intellect that given knowledge of all past and present facts, and the laws of nature, would be able to foresee the future down to the smallest detail. This idea was first given expression by the French mathematician and scientist Pierre Simon Laplace: “We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes” (1814, 4). Although this conception of determinism represents the traditional way of understanding the contemporary problem of free will and determinism, there have been other related and historically important threats. For example, divine foreknowledge or theological determinism has, for many, posed as much a threat to free will as natural laws. As Robert Kane writes: “Many theologians through the centuries have believed that God’s power, omniscience, and providence would be unacceptably compromised if one did not affirm that all events in the universe, including human choices and actions, were foreordained and foreknown by God. But many other theologians argued, with equal force, that if God did in fact foreordain or foreknow all human choices and actions, then no one could have chosen or acted differently, making it hard to see how humans could have ultimate control over their actions in a manner that would justify divine rewards and punishments. In such cases, the ultimate responsibility for good and evil deeds, and hence responsibility for evil, would devolve to God—an unacceptable consequence for traditional theists” (2002, 35).

**[B03n5]**

5. Another position similar to compatibilism but not mentioned here is *semi-compatibilism*. Semi-compatibilists maintain that moral responsibility is compatible with determinism but remain agnostic about whether free will is (see, for example, Fischer 1994; Fischer and Ravizza 1998).

**[B03n6]**

6. Bruce Waller maintains a skepticism of moral responsibility but not free will (see 2011 for details). Saul Smilansky’s position is also hard to place. While Smilansky maintains a skepticism about our purportedly commonplace belief in libertarian free will, and endorses the difficult insights of a hard determinist perspective, he also maintains that compatibilism retains some truth (see 2000; fn.12 below; and ch.6). Other recent books that advance skeptical posi-



tions, but are mainly written for a general public, include Harris (2012), Oerton (2012), Evatt (2010), and Pearce (2010).

7. . As quoted by Sommers (2007a, 61) and Strawson (ch.2).

[B03n7]  
[B03n8]

8. . Some skeptics, however, such as Benjamin Vilhauer (forthcoming), maintain an asymmetry in the justification of praising and blaming behavior according to which harmless praise can be justified in certain contexts but not blame.

9. . See, for example, Libet et al. (1983); Libet (1985, 1999); Soon et al. (2008); Wegner (2002); Wegner and Wheatley (1999); Bargh (1997, 2008); Bargh and Chartrand (1999); Bargh and Ferguson (2000); Wilson (2002); Nisbett and Wilson (1977); Doris (2002). The literature on *Social Intuitionism* (e.g., Haidt 2001) is also sometimes cited in this regard—see Sie (ch.16) for a brief discussion of its possible relevance. And for those unfamiliar with Wegner’s work, my reference here to the “double disassociation of the experience of conscious will” is to Wegner’s finding that the feeling of having willed an action can be doubly dissociated from actually having caused an action—that is, someone can experience themselves as having caused an action that they actually have not caused (e.g., I-Spy experiment), just as someone can think they have not caused an action that they actually have caused (e.g., alien hand syndrome, automatisms) (see Wegner 2002; Wegner and Wheatley 1999).

[B03n9]

10. . Predicting what revisions will be made is difficult. It’s possible that relinquishing the folk psychological idea of “soul control” will cause some to accept free will skepticism. But it’s also possible that some might adopt a *free-will-either-way* strategy causing them to accept compatibilism on pragmatic grounds, fearing the alternative.

[B03n10]

11. . *The Chronicle Review* (March 23, 2012) and *Scientific American* (April 6, 2010) respectively.

[B03n11]

12. . Smilansky’s *Fundamental Dualism*, however, also acknowledges that certain compatibilist insights are true. As Smilansky describes his position: “I agree with hard determinists that the absence of libertarian free will is a grave matter, which ought radically to change our understanding of ourselves, of morality, and of justice. But I also agree with the compatibilists that it makes sense to speak about ideas such as moral responsibility and desert, even without libertarian free will (and without recourse to a reductionist transformation of these notions along consequentialist lines). In a nutshell . . . ‘forms of life’ based on the compatibilist distinctions about control are possible and morally required, but are also superficial and deeply problematic in ethical and personal terms” (2000, 5; see also chapter 6).

[B03n12]

13. . According to Waller, “Blaming individuals and holding people morally responsible...is not an effective way of making either systems or people better; instead, it is a design for hiding small problems until they grow into larger ones and a design for concealing system shortcomings by blaming problems on individual failure. If we want to promote effective attention to the causes and correction of mistakes and the developments of more effective behavior and more reliable systems, then we must move away from the model of individual blame and instead encourage an open inquiry into mistakes and their causes and into how a system can be devised to prevent such mistakes and improve individual behavior” (2011, 291).

[B03n13]

14. . This descriptive thesis is only half of Nichols’ overall argument. In the second half of his paper he turns to the prescriptive question, *should* the acceptance of determinism lead to major changes in our lives? Here he argues that “there are good reasons to resist the cries for a revolution in our everyday lives” (2007, 406). This is what I have above labeled Nichols’ *anti-revolution* (see 2007 for details).

[B03n14]