Are We Free?

PSYCHOLOGY AND FREE WILL

Edited by

John Baer • James C. Kaufman • Roy F. Baumeister

Are We Free?

This page intentionally left blank

Are We Free?

Psychology and Free Will

Edited by
John Baer
James C. Kaufman
Roy F. Baumeister

OXFORD
UNIVERSITY PRESS
2008

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
www.oup.com

Library of Congress Cataloging-in-Publication Data
Baer, John.
Are we free? : psychology and free will /
John Baer, James C. Kaufman, Roy F. Baumeister.
p. cm.
Includes bibliographical references and index.
ISBN 978–0–19–518963–6
1. Free will and determinism.
I. Kaufman, James C.
II. Baumeister, Roy F.
III. Title.
BF621.B28 2008
123.'5—dc22 2007036151

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

For Sylvia

—JB

My work on this book is dedicated to my niece, Nicole Hendrix, who is living proof of the power of free will. Her perseverance, fierce intelligence, and independence will serve her well throughout her life. I am prouder than words can say.

Love,
JCK

This page intentionally left blank

CONTENTS

CONTRIBUTORS

David J. Atencio
Assistant Professor
Department of Individual, Family, and Community Education
University of New Mexico
Albuquerque, NM

John Baer
Professor
School of Education
Rider University
Lawrenceville, NJ

Albert Bandura
Professor
Department of Psychology
Stanford University
Stanford, CA

John A. Bargh
Professor
Department of Psychology
Yale University
New Haven, CT

Roy F. Baumeister
Professor
Department of Psychology
Florida State University
Tallahassee, FL

Daniel C. Dennett
Co-director
Center for Cognitive Studies
Professor
Tufts University
Medford, MA

Carol S. Dweck
Professor
Department of Psychology
Stanford University
Stanford, CA

Michael K. Goode
MA Student
Psychology Department
Washington University, St. Louis
St. Louis, MO

George S. Howard
Professor
Department of Psychology
University of Notre Dame
Notre Dame, IN

James C. Kaufman
Associate Professor
Department of Psychology
California State University
at San Bernardino
San Bernardino, CA

John F. Kihlstrom
Professor
Department of Psychology
University of California, Berkeley
Berkeley, CA

Alfred R. Mele
Professor
Department of Philosophy
Florida State University
Tallahassee, FL

William R. Miller
Professor Emeritus
Departments of Psychology
and Psychiatry
University of New Mexico
Albuquerque, NM

Daniel C. Molden
Professor
Department of Psychology
Northwestern University
Evanston, IL

David G. Myers
Professor
Department of Psychology
Hope College
Holland, MI

Shaun Nichols
Professor
Department of Philosophy
University of Arizona
Tucson, AZ

Steven Pinker
Professor
Department of Psychology
Harvard University
Cambridge, MA

Henry L. Roediger, III
Professor
Psychology Department
Washington University
in St. Louis
St. Louis, MO

Jonathan Schooler
Professor
Department of Psychology
University of California, Santa
Barbara
Santa Barbara, CA

Azim F. Shariff
PhD Student
Department of Psychology
University of British Columbia
Vancouver, BC Canada

Dean Keith Simonton
Professor
Department of Psychology
University of California, Davis
Davis, CA

Kathleen D. Vohs
Professor
Department of Marketing
University of Minnesota
Minneapolis, MN

Daniel M. Wegner
Professor
Department of Psychology
Harvard University
Cambridge, MA

Franklin M. Zaromb
PhD Student
Psychology Department
Washington University, St. Louis
St. Louis, MO

This page intentionally left blank

1 Introduction: Psychology and Free Will

John Baer
James C. Kaufman
Roy F. Baumeister

People generally act as if they possess free will, and they certainly act as though they believe in their own free will. People don't feel like automatons, and they don't treat one another as they might treat robots. And although people may acknowledge many external and internal factors that help shape their behavior and that of others, it seems that the buck must stop somewhere, and that somewhere is necessarily an important part of each person (whether one calls it a soul, or a personal identity, or a sense of personal responsibility). Humans may not be *totally* responsible for their behavior—if a gun is being held to someone's head and he or she is commanded to take certain actions, most people would agree that the person is not as responsible for those actions were there no gun and no command. In general, however, people implicitly assign a sense of agency, and of free will, to themselves and others.

Looked at closely, however, free will can be difficult to understand or to explain. Psychologists have tended to avoid the topic. As Roediger, Goode, and Zaromb write in their chapter of this book, the term *free will* didn't even merit an entry in the recent eight-volume *Encyclopedia of Psychology;* in fact, it didn't even appear in the index. Sometimes it's like a six-ton elephant sitting in the room, however—a thing very hard to ignore. In recent years, a number of psychologists have tried to solve one or more of the puzzles of free will (because free will raises not one, but many, tough questions). This book looks both at recent experimental and theoretical work directly related to free will and at ways psychologists deal with the philosophical problems long associated with the question of free will, such as the relationship between determinism and free will.

Does determinism rule out free will? On the surface, at least, it may seem to. But some philosophers have argued that determinism and free will are compatible. The problem may be that our intuitive concepts of free will simply don't make sense. Free will can't really mean that at any moment a person's behavior is totally unpredictable (and therefore entirely unconstrained). Such a universe would be, from psychology's perspective at least, the same as one governed entirely by chance, which is just another way of saying it is not governed at all. For psychology to make any sense, the universe must be, to some degree at least, predictable. A psychology that doesn't accept causes of behavior or the possibility of prediction is no psychology at all.

For those who accept free will as something real—whether that belief is based on determinism or not—how does free will work? What cognitive processes or mental structures underlie volition? What does it mean to choose, and how do people do it? And for those who believe it is an illusion, why does everyone believe in such an illusion? What evidence is there for either position? And is consciousness a requirement for free will? If so, how must we construe consciousness in order to understand free will? If conscious cognition is part of volition, but if (as some claim) that conscious cognition is completely determined by unconscious processes working in the background, does that still constitute conscious control of action? How can a psychology of conscious free will be tested and demonstrated experimentally?

It is the goal of this book to let psychologists from a variety of the discipline's subfields explain their beliefs about free will. Some of these psychologists are doing work that relates very directly to the questions raised by the puzzle of free will. Others do research that seems unrelated to questions about volition, and they therefore may not ordinarily write about free will, but they nonetheless think about it and about how our understanding of free will influences who we are. It has been our goal from the outset to include leading psychologists with a wide range of viewpoints, and we trust that readers will agree we have succeeded at least in doing that.

The question of free will is actually many questions, and the contributors to this volume have tried, in a variety of ways, to answer a variety of questions. We did not start with an outline or an agenda, but instead tried to include psychologists who come at free will from very different perspectives. Sometimes these perspectives are in direct disagreement, whereas in others they are simply addressing different questions. And sometimes there is even agreement.

We have book-ended the 15 chapters by psychologists and cognitive scientists with 2 chapters (2 and 18) that were written by philosophers. In chapter 2, Shaun Nichols suggests what philosophy might ask of psychology about free will. He outlines three distinct dimensions of the problem of free will about which psychology might make substantial contributions: a descriptive dimension that endeavors to discern the nature of lay views about free will, a substantive dimension that evaluates those lay views in light of what we know about psychological reality, and a prescriptive dimension that suggests how we should act in light of what we find out about the existence of free will. Alfred Mele's summary comes at the end, in chapter 18. In between these philosophical contributions, psychologists and cognitive scientists from various areas have tried to address the question of free will based on their understanding of both the psychological and the philosophical issues they find most significant.

The discussion among the psychologists starts with a chapter by David Myers, whose overview frames some of the key philosophical issues, such as the relationship of determinism and free will, in a way that will surprise some (e.g., "determinism encourages us to action, not resignation" and "determinism does not compel people to act against their will, nor does it deny them their experience of choice and their freedom to shape the future"). He addresses a broad range of questions, such as whether psychological science challenges or affirms free will; can we hold people accountable for their behavior if determinism is true?; and what psychology can say to ideas of free will as encoded in religious traditions. Because his chapter touches on so many topics that will be considered in more detail in the chapters to follow, we thought it would be a good place to start the conversation.

Carol Dweck and Daniel Molden begin their chapter by noting that the "nature of free will is ultimately a philosophical question; whether people *believe* they have free will is a psychological one, and whether people actually have free will is in the terrain somewhere in between." Much of the determinism–free will debate has been about how the laws of physics, not the laws of human nature that psychologists study, might constrain behavior. Dweck and Molden demonstrate how the *self-theories* that people have, and in particular individual beliefs about human qualities as either fixed or malleable, lead to different psychological realities, and they argue that incremental theorists have a stronger belief in free will than entity theorists.

Roy Baumeister suggests that the debate regarding the existence of free will may be an unproductive one and instead focuses on (a) how we might explain the common belief in free will and the phenomena to which that belief refers and (b) how free will might emerge and function, even in a psyche that is run largely via unconscious processes. He rejects the idea of free will based on randomness but argues that just as many philosophers accept some form of compatibilism that allows both free will and determinism, psychologists need not "fret that they will lose credibility as scientists if they, too, accept free will." He shows how evolution might have valued a conscious dispute-settling mechanism that could adjudicate among different unconsciously produced alternate actions and decisions, "possibly setting up and altering response tendencies that guide the automatic responses that are the immediate, proximal causes of behavior."

Albert Bandura argues that "metaphysical analytic preoccupation with the incompatibility of free will and determinism diverted attention from more fruitful analysis of the capacity of humans to bring their influence to bear on events," and he shows how thinking of free will in terms of the exercise of agency, which operates through a variety of cognitive and other self-regulatory processes, can help us understand how free choices are made. He concludes that nonagentic theories of behavior are simply a new incarnation of behaviorism, dismissing as they do such constructs as beliefs, goals, and expectations. It is precisely these (and other cognitive factors) that are needed for the cognitive control of behavior and for proactive moral agency.

John Bargh argues that historically, "free will has been the answer to the question of where our actions originate, where they come from in the first place. …But…there is no shortage of ideas or suggestions from our unconscious as to what to do in any given situation." Given the evidence for unconscious decision making from a variety of research paradigms, he suggests that we should begin with the assumption of mainly unconscious instead of conscious causation of action. His review of this evidence leads him to conclude that "there is no need to posit the existence of free will in order to explain the generation of behavioral impulses, and there is no need to posit free will in order to explain how those (unconscious) impulses are sorted out and integrated to produce human behavior and the other higher mental processes."

John Kihlstrom challenges the idea that we might be automatons in his chapter, "The Automaticity Juggernaut." He argues that although the cognitive revolution once again allowed the study of consciousness, the topic of consciousness continues to make many psychologists nervous and it is granted little causative power, reduced instead in many theories to an epiphenomenalist role that grants it no causal role in behavior at all. Like Bandura, he thinks this undoes the cognitive revolution and brings us full circle back to Skinner. Should we simply "jettison the notion of free will as a sentimental component of folk psychology that must be abandoned"? Or we might instead "accept the experience of conscious will as valid, and try to explain how free will can enter into the causal scheme of things in a material world of neurons, synapses, and neurotransmitters"?

Azim Shariff, Jonathan Schooler, and Kathleen Vohs argue that there are both easy and hard problems of free will, using as their model an argument from studies of consciousness, and suggest that most psychological research to date addresses what they call the easy problems. "The hard problem of free will represents *the* core problem of conscious free will: Does conscious volition impact the material world, and can phenomenal experiences translate into a physical events? And if so, how?" They review three main approaches to the hard problem—the hard determinist, compatibilist, and libertarian positions—and present their own recent research evidence which shows that when subjects are induced to believe that free will is illusory they behave less ethically, at least in laboratory experiments.

Henry Roediger, Michael Goode, and Franklin Zaromb note that although psychology may not be able answer the ultimate question (Does free will exist?), it has much to say about the control of behavior. They focus on four experimental research paradigms: Libet's response-choice paradigm, Logan's stop-signal paradigm, Jacoby's process-dissociation procedure, and Koriat and Morris Goldsmith's free and forced reporting procedure. These cognitive approaches together tell us a great deal about whether, and to what degree, human beings exercise control over their actions and decisions. Their focus is on finding ways to separate conscious and automatic influences on behavior to provide a window on volitional control. Much of our volitional control is more "free won't" than "free will," because it is rooted in our ability to inhibit unconscious responses. They conclude that experimental research can provide substantial evidence for partial conscious control of behavior while acknowledging that this is not the same as evidence of actual free will in the strongest sense: "We have danced around the issue of whether conscious control is to be equated with free will; in fact, we suspect that at the most basic level, the answer must be no. Even behavior that subjects believe to be completely under conscious control is influenced by external factors."

Daniel Wegner presents a series of studies that suggest that conscious will is an illusion that is rooted, in part, in our incomplete self-knowledge ("the mind presents us with only a relatively impoverished account of its own operations, and our attempt to make sense of the evidence yields the impression that we are freely willing our actions"). He argues that this is a powerful illusion—in fact, "the self seems remarkably resistant to reports of its demise, cropping up again and again in most every living human"—so powerful that even a scientist like himself who is steeped in the evidence that it is an illusion is "every bit as susceptible to the experience of conscious will as the next person." He compares the illusion of free will to visual illusions that we continue to experience even when we know they are illusions, and because of this persistence, he argues that we needn't fear grave social or personal consequences of scientific explanations that show free will to be illusory. Wegner closes his chapter by suggesting possible evolutionary scenarios that might have led to the evolution not of free will, but of the illusion of free will.

Daniel Dennett suggests that "so many really intelligent people write such ill-considered stuff when the topic is free will" because it matters so much, and "they just don't want to contemplate the implications straightforwardly." But often the real problem is that their ideas of free will are still tied to an outdated concept of free will that is rooted in Cartesian dualism. At the heart of the concept of free will, he argues, is the idea of moral responsibility, and a thoroughly materialistic understanding of free will—without the illusion of "the inner puppeteer who pulls the body's strings"—can find all the free will we need, "distributing its tasks throughout not just the entire brain, but the body and the 'surrounding' cultural storehouse." He notes that recent research raises important questions about the impact of holding varying beliefs in free will. He concludes that we "need to coordinate our investigations of the role of censure and punishment…with our investigations of the complexities of human motivation, and the role of beliefs—and beliefs in beliefs" to understand how new conceptions of free will might influence behavior and our sense of self and responsibility for our actions. "This is going to be a ticklish task, in which missteps might be painfully amplified. No wonder our hands shake when we get to work on it."

George Howard also argues that the free will–determinism debate has been hindered by the way it has been framed. There are two related but different questions, or dimensions: the power of self-determination versus mechanistic or nonagentic determination, and complete determinism versus complete acausality. Regarding the latter, "If you want to be a scientist, you had better be a determinist." But the former dichotomy, between self-determination and mechanistic determination, is a false one. A psychologist can believe in both. He reports a series of studies that measure degrees of self-control, showing that although in some areas "the amount of control they exhibit is vanishingly small," in others it is "enormous." He concludes that human behavior is partially self-determined and partially nonagentically controlled.

Like Howard, William Miller and David Atencio believe that questions of free will have "often been cast as a dichotomous choice between free will and determinism," but in fact neither extreme view can be correct. They propose ways to measure what they call the "*volitionality* of behavior," the degree to which some behavior is subject to willful control. Some behaviors have higher volitionality, and some people have greater degrees of volitional control. They conclude that psychology must find ways to understand volition as a significant determinant of both individual and group behavior.

Dean Simonton, one of the world's leading creativity researchers, notes that "human creativity represents something of a paradox" because "few areas of human behavior require so much will power" and yet in "few areas is the will so powerless." Creativity seems to be determined both by outside forces beyond the individual's control, but at the same time the very act of creation seems like more than anything an act of will. This is especially true of creativity at the highest level.

In the final two chapters (prior to Alfred Mele's summary and conclusions), John Baer and Steven Pinker examine specifically the connection between determinism and free will. Baer argues that, rather than choosing between determinism and free will, free will in fact requires determinism (as does psychology). To whatever extent behavior is due to chance, free will cannot exist; but a determinism that includes effects of differences in personality, cognitive abilities, beliefs, ideas, emotions, memories, wishes, and thinking styles on volition makes possible the kind of free will that most of us believe intuitively that we possess. Pinker confronts several fears that determinism engenders, such as the fear that "deep down we are not in control of our own choices" and the fear that determinism makes it impossible to hold anyone accountable for their actions. Although the fear of biological determinism seems to many more frightening, environmental determinism must carry the same baggage. But "contrary to what is implied by critics of biological *and* environmental theories of the causes of behavior, to explain behavior is not to exonerate the behaver." Pinker explains why this is so, concluding that "I do not claim to have solved the problem of free will, only to have shown that we don't need to solve it to preserve personal responsibility in the face of an increasing understanding of the causes of behavior."

In the final chapter of the book, Alfred Mele has distilled what the various contributors have told us about free will. It is impossible to summarize in one paragraph Mele's commentary because his analysis covers so much territory—he is, after all, commenting on the full range of ideas presented elsewhere in the book. Among the points he makes, however, are some that tend to find agreement between viewpoints that were seemingly at odds. For example, differing views on the role of determinism in both psychology and free will may be rooted in different definitions of determinism; by clarifying definitions, some of the disagreements fall away. He also examines the connection between consciousness and free will and the implications of Libet's studies, which many of the chapters discuss. He shows why this paradigm may have much less to say about free will than often claimed, and he also proposes some new studies that psychologists might wish to consider. He concludes that "One may find that some of the conceptions [that people have of free will] are self-contradictory, that others are hopelessly magical or mysterious, and that yet others suggest potentially fruitful research programs. One would expect most scientists with an experimental interest in free will to be attracted to conceptions of the third kind."

There is, of course, no single and irrefutable answer to the many questions posed by free will. There is, however, evidence that makes possible both a better understanding of what free will is or might be and the construction of more psychologically sound theories of free will. We believe readers will find that the contributors to this book have made huge progress in defining key issues, marshaling relevant research findings, and explaining what psychology can contribute to this important conversation.

2 How Can Psychology Contribute to the Free Will Debate?

Shaun Nichols

BACKGROUND: THREE PROJECTS

Are people free and morally responsible? Or are their actions determined, in other words, inevitable outcomes of the past conditions and the laws of nature? These seem to be fairly straightforward questions, but it is important to distinguish three different dimensions of the free will debate: a descriptive project, a substantive project, and a prescriptive project. In this chapter, I'll consider how psychology can contribute to each project in turn. First, I should say a bit more about the projects.

The goal of the descriptive project is to determine the character of folk intuitions surrounding agency and responsibility. By uncovering the folk intuitions, one hopes to be able to sketch out the folk theory that underlies these intuitions.[1] Of particular interest for the free will debate is whether the folk notions of choice and moral responsibility are consistent with determinism. *Incompatibilists* maintain that our conceptions of free will and moral responsibility are at odds with determinism. *Compatibilists* deny this and insist that our notions of free will and moral responsibility are consistent with determinism.[2]

The goal of the substantive project is then to determine whether the folk views are correct. Given the folk concepts and the way the world is, does free will exist? Are people morally responsible? *Libertarians* maintain that we do have indeterminist free will (e.g., Campbell 1957, Kane 1996, O'Connor 1995). *Eliminativists* about free will maintain that free will doesn't exist. The best known version of this view is "hard determinism," according to which we lack free will because determinism is true. However, many free will eliminativists maintain that even if determinism is false, we still lack the kind of indeterminist choice that is required by the folk notion (e.g., Pereboom 2001, Sommers 2005, Strawson 1986). On this view, our notion of free choice is incompatible with the facts, regardless of whether determinism is true or false.[3]

The prescriptive project is different from both the descriptive and substantive projects. Here the question is how we should react, given what we know about our concepts and the world. Should we revise or preserve our practices that presuppose moral responsibility, like practices of blame, praise, and retributive punishment? It might be helpful to summarize with a tree diagram the philosophical positions that emerge from these questions (see figure).

Let's start at the top with the central descriptive question, "Is the folk concept of free choice compatible with determinism?" If the answer is "yes," then *compatibilism* is the right view, and at that point, at least as far as the free will/determinism issue goes, we need not bother with the substantive and prescriptive matters. If our concept of free will is happily consistent with determinism,

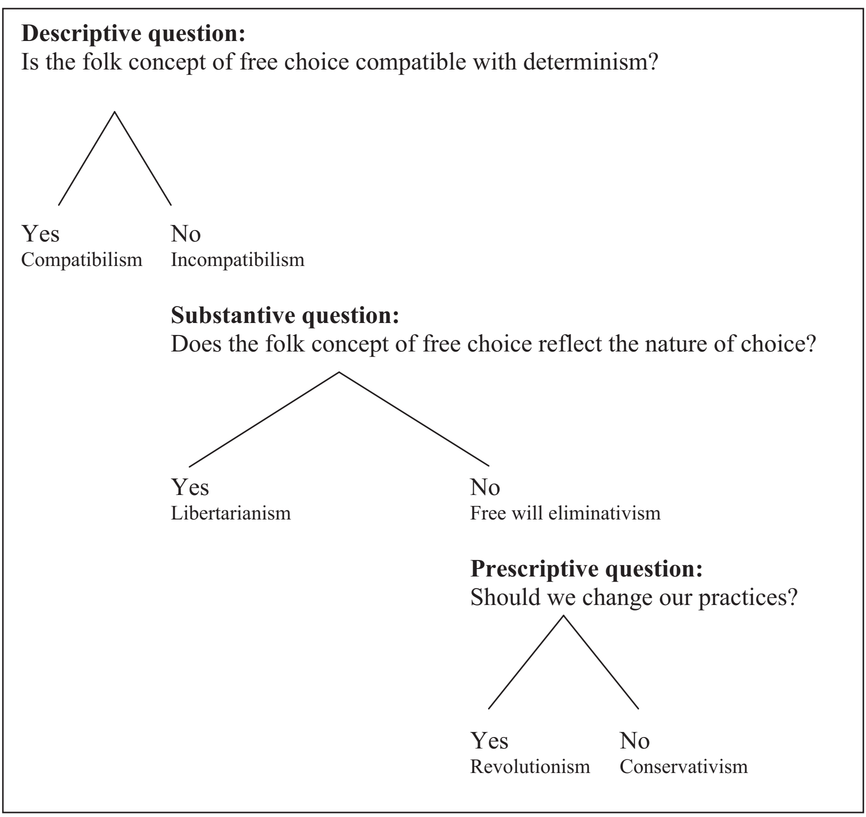

Figure 2.1.

then determinism poses neither a substantive nor a moral threat to our current views and practices (Hume 1955/1743).

However, if the answer at this juncture is “no,” then *incompatibilism* is the right view (Kant 1956/1788, Reid 1969/1788, D’Holbach 1970/1770). If so, we face a pressing substantive question: “Does the folk concept of free choice reflect the nature of choice?” In particular, it becomes a major concern whether determinism is true. *Libertarianism* maintains that choices are *not* determined. Libertarians typically hold that on the folk view, choices are not determined, and also that the folk view of choice matches the way the world is. If libertarianism is right then the prescriptive question is not pressing. Our normal practices of regarding people as free and responsible are perfectly appropriate—people *are* free and responsible.

According to *free will eliminativists*, however, the answer to the substantive question is that we lack the kind of free will we think we have. On their view, the facts about the world are at odds with the way we think of ourselves. If this worrying view is correct, then the prescriptive question takes on great significance. If our folk view of choice is wrong, what is the appropriate response? Should we stop treating each other as free and morally responsible agents? *Revolutionism* is the view that we should overhaul our practices that presuppose free will and moral responsibility. *Conservatism* is the view that we should leave practices more or less untouched. Now let’s see what psychology can do for each of these projects.

THE DESCRIPTIVE PROJECT

Many philosophical issues are resolutely technical and detached from lay commitments. If you tell undergraduates that some philosophers, the logicists, think that mathematics can be derived from logic, the typical response will be “Who cares?” Most people simply don’t have intuitions about the logical foundations of mathematics. This is not what happens when students are told that some philosophers maintain that every decision a person makes is an inevitable consequence of what happened prior to the decision. That is a disturbing suggestion even to students whose only interest in logicism is whether it will be on the final exam. The problem of determinism and free will strikes a deep worry in us. The fact that the problem of free will resonates with people is a *psychological* fact. The descriptive project strives to capture the nature of our response to the problem of free will—what is it that we think about free will and responsibility? Is our notion of free choice incompatible with determinism? Where does our notion of free will come from? Psychology is obviously in an excellent position to help answer these questions.

The Folk Notion of Choice

Do the folk have libertarian views about choice? This issue has only recently been explored empirically. Over the last few years, there has been a modicum of evidence that we do have indeterminist views about choice. Recent experiments investigated whether children think that agents could have done otherwise than they did (Nichols 2004a). Children were placed in one of two conditions: those in the *agent* condition witnessed an agent exhibit motor behavior; those in the *object* condition witnessed an object move. For instance, children in the agent condition were shown a closed box with a sliding lid; the experimenter slid the lid open and touched the bottom. Children in the object condition were shown the closed box with a ball resting on the lid; the experimenter slid the lid open and the ball fell to the bottom. The child was asked whether the agent/object had to behave as it did after the lid was open, or whether it could have done something else instead.

The results were very clear. Every single child said that the person could have done something else, and nearly every child rejected this option for the object. In a second study, adults and children were asked about physical events, for example, a pot of water coming to a boil, and moral choice events, for example, a girl stealing a candy bar. Participants were asked whether if everything in the world was the same up until the event occurred, the event had to occur. In this setting, both adults and children were more likely to say that the physical events had to occur than that the moral choice events had to occur. This provides preliminary evidence that the folk have a concept of free choice on which agents *could have done otherwise*.

Further support for the claim that people regard choice as indeterminist comes from recent experiments that Joshua Knobe and I have conducted (Nichols & Knobe 2007). We presented subjects with a questionnaire that depicted both a determinist universe (A) and an indeterminist universe (B), described as follows:

> The key difference…is that in Universe A every decision is completely caused by what happened before the decision—given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does **not** have to happen* the way that it does.

After this description, subjects were asked, "Which of these universes do you think is most like ours?" The vast majority of subjects answered that the *indeterminist* universe (Universe B) is most like ours. Note that the only feature of the universe that is indeterminist is choice. So, the responses indicate that people are committed precisely to the idea that choice is indeterminist.[4]

The above work represents merely one view (for a contrasting account, see Nahmias et al. 2006; Nahmias forthcoming). Obviously further work would need to be done to confirm that people have indeterminist views about decisions. Even apart from this controversy, there are important additional questions that are ripe for psychological exploration. Perhaps the most interesting question is whether the results on folk indeterminism would extend to other cultures. In light of the findings on cross-cultural differences in intuitions and attitudes (see, e.g., Nisbett et al. 2001; Machery et al. 2004), it seems quite possible that the preliminary findings of indeterminist views about free choice will be restricted to Western culture. On the other hand, it might be that the belief in indeterminist free will is, as some philosophers have maintained, a deep feature of creatures like us (see, e.g., Strawson 1986; Nagel 1986).

Where Does the Belief in Libertarian Free Will Come From?

Assuming we have a notion of libertarian free will, a new question emerges for the descriptive project: How did this notion come about? Several accounts of the origin of our belief in libertarian free will have been offered. None of the accounts, however, has achieved adequate support. I'll review some of these accounts and their evidentiary shortcomings, then I'll suggest a different, but also unsupported, account of the belief.

The traditional explanation for how we come to believe in indeterminist agency is that it comes from introspection (e.g., Reid 1969/1788, Holbach 1970/1770) because introspection fails to reveal any deterministic underpinnings of my decision making. Of course, both libertarians and determinists can agree that introspection fails to reveal deterministic decision making. The determinist maintains that we fail to introspect the deterministic processes that actually produce our behavior. The libertarian will insist that there is no such deterministic process, so of course we don't introspect it.

Although the fact that we don't introspect deterministic causes of our choices is almost certainly part of the story, it can hardly be a complete explanation of how we acquire the belief in indeterminist choice. The fact that we don't perceive deterministic decision-making processes doesn't yet explain why we would believe that our decisions are *indeterministically* generated. After all, we often think processes are deterministic even when we don't perceive deterministic causal transactions.[5] Even some of our behaviors have this quality. When my eye twitches, I have no idea what causes it, but this doesn't remotely lead me to think that eye twitches are indeterministically generated. Thus the fact that we don't perceive a deterministic process of decision making must be supplemented to explain the intuition that our decisions are not determined. The natural supplement would be to maintain that we have a standing belief that we do have introspective access to all the causal processes underlying our own decision making. If we have a standing belief in such introspective transparency and we fail to introspect deterministic processes, we might infer that there are none.

This combination of presumed introspective access coupled with introspective failure might provide an explanation for how we come to believe in our own indeterminist agency. But there are several shortcomings. First, we would need evidence that people do in fact have a standing belief that we have introspective access to all the causal processes underlying our own decision making. Second, the account also has to assume that people carry out the required inference to arrive at the view that agency is indeterminist. Third, this account would apply only to oneself. It would need to be supplemented to explain why we think that *other* people's decisions are indeterministic. None of these shortcomings is decisive, of course. But clearly psychology has an important role to play in investigating the promise of this account.

Another tempting account of the genesis of the belief in free will is that it is inferred from seeing creatures exhibit spontaneous motion. That might spawn the idea of freely generated action. Joshua Greene and Jonathan Cohen (2004) make some suggestive remarks along these lines. First, they note, in keeping with much work in developmental psychology (e.g., Baron-Cohen 1995; Leslie 1995), that the mind plausibly has different systems for dealing with matter and for dealing with minds. The system for dealing with minds is triggered by a wide variety of stimuli, including the Heider/Simmel animation of spontaneously moving geometric shapes (e.g., Heider & Simmel 1944). Greene and Cohen link their proposal to such findings and maintain that we regard others as free because we regard them as having minds. They write, "we suggest that a crucial feature, if not the defining feature, of the mind (intuitively understood) is that it's an *uncaused causer* [italics added] (Scholl & Tremoulet, 2000). Minds animate material bodies, allowing them to move without any apparent physical cause and in pursuit of goals" (p. 35). On their view, ordinary objects like rocks seem to obey ordinary physical laws—"these things don't get up and move around on their own." There are other things however that "seem to operate by some kind of magic . . . [moving] about at will, in apparent defiance of the physical laws that govern ordinary matter" (p. 32). This suggests the following sort of account: Spontaneous motion triggers the attribution of a mind as a free agent—an uncaused causer. Although Greene and Cohen don't endorse quite this picture, it is an attractively simple hypothesis about why we believe in free will.

Attractive, but mistaken. First, it is quite possible that the system for dealing with predicting and explaining minds is *deterministic*. Indeed, in the vast literature on "mind reading," the notion of indeterminist choice is not invoked in any of the prevailing models of how children predict and explain behavior (see, e.g., Gopnik & Meltzoff 1997, Leslie 1995, Nichols & Stich 2003). That is, none of these models maintains that people routinely invoke the notion of indeterminist choice when they predict and explain behavior. This makes sense, because the predictive aim of the mind-reading system would not be improved by the inclusion of an assumption of metaphysical indeterminism. So, we cannot assume that we get attributions of libertarian free will as a freebie once we have attributions of minds.

Second, even babies apparently attribute mental states to various phenomena, including certain computer-generated geometric objects of the Heider/Simmel sort adverted to by Greene and Cohen (e.g., Csibra 2003; Kuhlmeier et al. 2003; Premack 1990). Recall that to attribute libertarian free will to an agent is to maintain that the agent could have done otherwise even if everything else had been the same; that seems a more sophisticated thought than many of us are willing to impose on the baby.

Finally, recent evidence from Susan Johnson and colleagues suggests that spontaneous motion is not sufficient to activate attributions of mental states.[6] In one experiment, 12-month-old infants were shown a fuzzy brown object under a variety of different conditions (Johnson, Slaughter, & Carey 1998). In one condition, the fuzzy brown object interacted contingently with the infant by beeping and flashing when the infant babbled or moved; in another condition, the fuzzy brown object exhibited an equivalent amount of apparently self-generated flashing and beeping, but in this condition the activity was not contingent on the infant's behavior. In both conditions, children's looking behavior was measured when the fuzzy brown object "gazed" at one of two objects by making a smooth, 45-degree turn toward the object and remaining in this orientation for several seconds. What Johnson and colleagues found was that infants would follow the "gaze" of the fuzzy brown object when its spontaneous activity was contingent, but not when the spontaneous activity was noncontingent. Johnson and colleagues propose that what happened in the experiment is that the infants followed the gaze when the fuzzy brown object was coded as an *agent*. Indeed, gaze following is often taken to reflect the "implicit attribution of a mind to the gazer" (Johnson et al. 1998, 233).[7] This suggests that mere self-generation of behavior isn't sufficient for attribution of a mind because in the noncontingent condition, infants don't seem to attribute mental states despite the presence of spontaneous motion. Of course, it's possible that babies do attribute minds and free will to these spontaneous movers, and then quickly revise their beliefs. But without evidence to that effect, the spontaneous motion account does not yet explain the attribution of free will.

In Nichols (2004a), I suggested, *very tentatively,* an alternative account on which the acquisition of the belief in indeterminist choice derives from a prior belief in obligation. According to a famous Kantian argument, we can prove that we have indeterminist choice from the maxim "*ought* implies *can*" and the fact that we ought to follow the moral law. The idea is that we can't be obligated to do the impossible, and if determinism is true, it is impossible for us ever to do other than what we are determined to do. Thus, if we say that a person *ought* to have behaved differently, this implies that the person *could have done otherwise* (in an indeterminist sense). The suggestion in Nichols (2004a) was that, despite the dubiousness of the Kantian argument as a *proof* of indeterminist choice, it might provide an account of how we come to *believe in* indeterminist choice. There is plenty of evidence that even young children think that people *ought* to behave in certain ways (e.g., Nichols 2004b, Nucci 2001). Indeed, the child applies notions of obligation in a variety of contexts including contexts of moral transgressions (you shouldn't kick people), conventional transgressions (you shouldn't eat steak with your hands), and even simple cases of advice (you should put on sunscreen). If children apply some notion of obligation that carries the Kantian implication *could have done otherwise* (in an indeterminist sense), then the child has the essential ingredients for coming to believe that decisions are not determined.

Unfortunately, although there is abundant research showing that children apply obligation concepts, there is no evidence yet confirming the idea that children embrace the Kantian maxim. It is likely that children embrace *some* kind of ought-implies-can view. If you ask whether it was wrong for the paraplegic not to swim to save a drowning victim, children will presumably say that it's not wrong because he *couldn't* swim. But it will be harder to show that children think that obligations carry the implication of *indeterminist*-can, and that this leads to their belief in libertarian free will.

My own newly favored view about acquisition is that the notion of indeterminism plays a larger role in childhood cognition than previously recognized (Nichols 2006). Even apart from choice events, the child often hears and uses a notion of possibility that is naturally interpreted as in conflict with determinism. Here are some examples drawn from the CHILDES database in which children advert to possibilities:[8]

FATHER: You could fall and get hurt Ross.
ROSS (4 YEARS AND 2 MONTHS [4;2]): No. Not if I hold on to here and here I won't.
FATHER: You could…It's dangerous (MacWhinney & Snow 1990).

ROSS (2;7): Marky [a younger sibling] might fall (MacWhinney & Snow 1990).

ADAM (4;2): Paul [a younger sibling], you might cut yourself on this (Brown 1973).

It's natural to interpret these uses of *possibility* as in conflict with determinism. That is, it's natural to read Ross as saying that it's a genuinely open possibility that Marky will fall (or not fall), and Adam as saying that it's a genuine possibility that Paul will cut himself (or not). There are various determinist-friendly ways to read the language of possibility, and perhaps one of these ways is the best interpretation of children's modal language. But it's worth noting that some of most familiar ways to tame modality in philosophy don't look at all promising.

One way to preserve a notion of *possibility* while being neutral about determinism is to treat *possibility* as a deflationary kind of epistemic possibility, so that when I say "*p* is possible," what that really means is *For all I know, p will happen*. But this seems an implausible interpretation. When Adam says that Paul might cut himself, it seems unprincipled to maintain that Adam really just means, "As far as I know, Paul will cut himself." Similarly, when Ross' father says, "You could fall," and then repeats, "You could…It's dangerous," it's doubtful that Ross would interpret his dad as merely reporting on epistemic possibility. The simple epistemic possibility interpretation is even less plausible when we move to statements about past possibilities. Parents say things like, "You could have broken the lamp!" And kids come to use language this way as well. In CHILDES we find Ross (at age 5) saying that he climbed on a shelf and "It could have fell on us" (MacWhinney & Snow 1990). Obviously he doesn't mean, "For all I know, it fell on us." Thus, the simple epistemic gloss fails to provide a general account of children's judgments of possibility.

If the child does have an indeterminist notion of *possibility* outside of choice contexts, we get a new avenue for exploring the acquisition of the concept of free will. For now it seems that the child's indeterminism about choice is part of a more general indeterminism about possibilities. Elsewhere I've suggested that a primary function of the modal concept POSSIBLE is to represent risk and opportunity (Nichols 2006b). This was based on an informal review of modal talk in the CHILDES database. It is a salient fact about everyday conversation that children and parents tend to use modal language largely to convey information about risks and opportunities. We've already seen some examples that represent risks. Here are some examples of children deploying modals to represent opportunities:

ALISON (2;4): We could march around or we could run around (Bloom 1973).

ROSS (3;3): Hi Titus [a cat]. I got her tail.
FATHER: You did.
ROSS: She's under there. I could get her.
FATHER: Okay but don't be too mean to her okay? (MacWhinney & Snow 1990).

FATHER: I can't make that one work.
ABE (3;7): You could glue it (Kuczaj & Maratsos 1975).

ADAM (4;1): We could put the animals in here (Brown 1973).
ADAM (4;10): Ursula, you could stay and eat with us (Brown 1973).

And here are remarks about opportunities made to children by adults:

MOTHER: Adam, you could draw some pictures now (Brown 1973).

MOTHER: You could shoot at anything that's set up for a target (Brown 1973).

FATHER: You could go visiting to other people's houses. That'd be fun (MacWhinney & Snow 1990).

MOTHER: You could play with your birdies honey (Sachs 1983).

As parents, when we point out risks to our children, what matters to us is making sure that our children are safe, that they avoid unnecessary dangers. We typically don't care about trying to communicate a carefully qualified notion of possibility that is neutral about determinism. Similarly, when we point out opportunities to our children, we aren't concerned to get them to hedge the modal notions in a compatibilist way. We want to stress the options before them. In short, when we alert our children to risks and opportunities, deterministic explanation is pretty much the last thing on our minds. Or theirs. In this light, it should not be surprising if our notion of possibility fails to be nuanced in a compatibilist fashion. The concept of *possibility* can serve the key function of representing risks and opportunities quite well without any compatibilist subtlety.[9]

Although the above account seems most plausible to me at the moment, the lay understanding of modality is dramatically underexplored. In particular, there is precious little experimental evidence on the extent to which children apply an indeterminist notion of *possibility.* Once again, psychology is uniquely well positioned to confirm or disconfirm the acquisition story.

THE SUBSTANTIVE PROJECT

Substantive questions about free will and determinism—are our choices determined?—are most pressing for incompatibilists. If free will is incompatible with determinism, then it is imperative that we know whether our actions are determined. The arguments in favor of libertarian free will typically appeal to introspection (e.g., O'Connor 2000, Campbell 1957). Because few psychologists would countenance such arguments, we will consider here only arguments against libertarianism.

A Priori Arguments

A number of philosophers argue on a priori grounds that libertarian free will is incoherent or impossible (e.g., Double 1991; Strawson 1986, 1994). Perhaps the most enduring a priori critique is Hobbes's libertarian dilemma. On the one hand, libertarians say that an agent's decision isn't free if the decision is determined. But on the other hand, if the decision is *not* determined, then it isn't determined by the agent either! That, critics maintain, leaves the libertarian fresh out of intelligible options (for discussion, see Kane 1996, p. 11). As the libertarian dilemma illustrates, libertarian free will can seem decidedly mysterious and counterintuitive, and this counts as a strike against it. But, as evidenced by quantum mechanics, some mysterious and counterintuitive views are apparently true. So it would be overly hasty to conclude from such a priori arguments that libertarian free will is metaphysically impossible.

A Posteriori Arguments

In any case, whatever one wants to say about the a priori arguments, this is not a natural place for psychology to make a contribution. A more natural role for psychology is in making a posteriori objections to libertarian free will. Psychology might, for instance, show that (psychological) determinism is true, thereby dealing a direct blow to libertarianism. The most explicit development of this kind of attack comes from John Bargh and Melissa Ferguson (2000). Their argument runs into serious troubles, I think, and the troubles illustrate why it will be very difficult for 21st-century psychology to prove that libertarian free will doesn't exist. So I want to discuss Bargh and Ferguson's argument at some length.

Bargh and Ferguson adopt a standard notion of determinism: "For every psychological effect (e.g., behavior, emotion, judgment, memory, perception), there exists a set of causes, or antecedent conditions, that uniquely lead to that effect" (2000, p. 925), and they claim that psychology has provided evidence for determinism. For instance, they write, "Behaviorists and cognitive (and social-cognitive) scientists have accumulated evidence of determinism by their many demonstrations of mental and behavioral processes that can proceed without the intervention of conscious deliberation and choice" (p. 925). Bargh and Ferguson rightly note that the existence of consciously controlled decisions does nothing to refute determinism. But they then go on to argue for the much stronger thesis that work on control processes provides clear evidence for determinism: "The automatic goal operation experiments provide…rather obvious evidence that even controlled mental processes are themselves controlled and determined" (p. 939).

As an example of such evidence, Bargh and Ferguson advert to an important experiment by Chartrand and Bargh (1996). In this experiment, participants first completed a scrambled sentence task in which they are told to form grammatical sentences out of short lists of words. In one condition (impression goal condition), the task was loaded with terms associated with impression formation (e.g., "evaluate," "judge," "assess"); in the other condition (memory goal condition), the task was loaded with terms associated with memory (e.g., "retain," "remember"). All participants then read several sentences describing the behavior of a given person and were told that they would be asked about this later.

Strikingly, participants in the impression goal condition had more accurate recall than the other participants. These participants also showed better organization of the information in memory. This effect conforms to earlier results in which subjects are explicitly instructed to inform an impression or to memorize. What Chartrand and Bargh show is that even when the goal is induced implicitly, the effect still emerges. Indeed, in their experiment, participants in the two conditions were unaware that the goal had been induced—the two groups showed no difference when asked what they were trying to do when they read the descriptions of the person's behavior (Chartrand & Bargh 1996).

There are a number of problems with using this evidence to try to undermine libertarian free will. First, libertarians maintain that determinism is false about *decisions*. Libertarians have typically been willing to allow that determinism is true about other psychological processes. Libertarians take considered *moral* decisions to be a paradigm example of the kind of mental activity that is a good candidate for not being determined. Strategy formation of the sort demonstrated by Chartrand and Bargh does not fit this paradigm very closely.

Although this might be a problem with using this particular bit of evidence to support determinism, I suspect that the ever ingenious John Bargh could generate evidence that even moral decisions are affected by nonconscious goals. But there is a deeper with problem with the argument. The results are, of course, statistical. What they show is that nonconscious goals *influence* psychological outcomes. That's a far cry from showing that the psychological outcomes are *determined*.

To make this point as plain as possible, consider another delightful result from Bargh's lab (Bargh, Chen, & Burrows 1996). Again, subjects were given a scrambled sentence task. This time participants were assigned to different conditions in which the word sets were loaded with terms related to rudeness, to politeness, or to neither. All participants were told that after completing the scrambled sentence task, they were to go tell the experimenter that they were ready for the next task. However, the experimenter was engaged in conversation, and the participant would have to interrupt in order to tell the experimenter. Bargh and colleagues found that among those primed for rudeness, 63% interrupted, among those primed for politeness, only 17% interrupted, and for the control group, 37% interrupted. Again, subjects didn't subsequently explain their behavior by invoking exposure to terms of politeness or rudeness. This is a remarkable demonstration that our decisions are influenced in ways that fall well outside of our awareness. But obviously there's still lots of individual variation here. The politeness prime didn't eliminate interruptions, and the rudeness prime didn't send interruptions to ceiling. A libertarian can perfectly well maintain that indeterminist free choice accounts for some of the variance in these studies.

A more general point can be made now. Virtually all libertarians are happy to allow that there are many factors that influence our choices (Campbell 1957, Clarke 1993, Kane 1996, O'Connor 1995). Our natural inclinations, genetic predispositions, and involuntary appraisals all influence outcomes. The terrific work in social psychology shows that there are hitherto unrecognized influences. But they are still just more factors that the libertarian can happily concede. The difficulty in excluding indeterminist choice is that one would need evidence of something stronger than influence. We would need to show that the decision was *entirely* produced by a known set of influences. As far as I know, we don't have a single worked out case of the deterministic processes that generated a single decision of a single individual. Thus, we are not in a good position to claim that determinism has been demonstrated.

The challenge for the a posteriorist is actually even greater because even if we found that many decisions are determined, this still wouldn't refute libertarianism. A libertarian can perfectly well maintain that libertarian free activity is a relatively rare phenomenon. Indeed, one prominent scientific libertarian, Robert Kane, maintains that even my "free" acts can be determined, so long as certain determining factors have their ultimate source in a "self-forming" event. Even though they might be rare, these self-forming events are the key to our ultimate freedom and responsibility according to Kane (1996, pp. 75–78). Put simply, the problem for a posteriori arguments for determinism is that determinism is a universal claim—every event is determined. Libertarians, by contrast, don't think that every event is indeterministic. Indeed, some libertarians don't even think that every choice event is indeterministic. Thus, to establish determinism on the basis of psychological evidence would require evidence that suffices for a universal generalization. That's a towering order. At best, it will be a long time before the psychological sciences exclude the rare occurrence of indeterminist free will.

Although psychology is not currently equipped to prove that determinism is true, it is plausible that determinism is an important guiding principle for psychological research. This is also part of Bargh and Ferguson's brief. They write, "Psychologists studying higher mental processes should continue the scientific study of conscious processes but at the same time give appropriate attention to the deterministic philosophy that must underlie such analysis" (2000, p. 940). Here, they are on much better ground. I do think that psychological determinism has been and will continue to be a vital assumption guiding research. And I'm inclined to think it's true. But my allegiance was not produced by checking the evidence. Rather, it came from an abiding conviction that people's decisions *have* to have an explanation.

Genetic Arguments

Even if the a priori and a posteriori arguments against libertarian free will are limited, there is another way that psychology can contribute to the substantive project. Rather than arguing that libertarian free will doesn't exist, we might argue that our belief in libertarian free will is unjustified. Freud used genetic considerations to argue that religious belief is unjustified. On his view, when we see the source of our religious beliefs, we will come to appreciate that they are unwarranted (Freud 1961/1927). Similarly, we might find that the source of our belief in libertarian free will reveals that the belief is unwarranted.

In the section "Where Does the Belief in Libertarian Free Will Come From?," I reviewed several different accounts of how we come to believe in libertarian free will, and at least in some cases, if the accounts are right, then it seems that we are not justified in our belief in libertarian free will. To see this, let's consider the last account that I reviewed in that section, coupled with the idea that we lack introspective access to any deterministic processes. I suggested that our belief in libertarian free will is part of a more general indeterministic outlook that applies to risks and opportunities. Parents present risks and opportunities as genuine, indeterminist, possibilities, and children (and adults) represent them as such. Because the purpose of these representations is to avoid danger and capitalize on opportunity, there is no direct cost in failing to hedge them in compatibilist ways. The representations function equally well regardless of whether they accurately represent indeterminist events or whether they inaccurately represent events that are deterministic but unpredictable by us.

So are these indeterminist beliefs warranted? In the case of our indeterminist beliefs about risks (e.g., "The branch might fall on you!"), the right thing to say is that our beliefs are unwarranted. We lack the discrimination to see whether natural events concerning middle-sized objects are deterministic. Indeed, our everyday observations are entirely consistent with a deterministic physics, even if more careful observation exposes indeterminism. Because of our limited powers of discrimination in everyday life, we would believe events to be indeterministic even if they aren't. As a result, when the child (or adult) says, "The branch might fall on you," she isn't warranted in thinking that it's a *metaphysically* indeterminist possibility. Rather, she is warranted in a belief only about *epistemic* possibility—*for all I know the branch will fall.*

Similarly, we lack the discrimination to see whether the choices people make have a libertarian source. Obviously, we lack the necessary discrimination when it comes to other people—we have no direct access to the psychological processes that eventuate in their decisions. But research in social psychology suggests that even introspection fails to provide the kind of access that would be needed to detect whether our own choices are generated by libertarian free will because introspection fails to detect the causal influence of many causal factors, including both conscious and unconscious mental states (e.g., Bargh & Chartrand 1999, Nisbett & Wilson 1977, Wegner 2002, Wilson 2002). As a result, even if our decisions are deterministically generated, introspection would still fail to reveal this to us. So, our belief in libertarian free will depends crucially on mechanisms that are too insensitive to tell us whether our choices are in fact generated by indeterminist free will. Prima facie, then, our belief in libertarian free will lacks any good rational foundation. Note that this is not to say that the evidence on introspection shows that we lack libertarian free will; rather, the evidence helps to show that our belief in libertarian free will is not well grounded. Thus, although we might not have compelling empirical evidence against libertarian free will, we lack good reasons to trust our belief in libertarian free will.

It's useful to contrast this epistemic situation with our position with respect to quantum mechanics. As noted in the section "A Priori Arguments," libertarian free will is weird—it's hard to devise a clearly intelligible and intuitive account of libertarian free will. But the same can be said for quantum mechanics. Quantum mechanics is so deeply weird that there's some question about whether we have the cognitive resources to understand it. So what's the difference between quantum mechanics and libertarian free will? The difference is that we came to believe in quantum mechanics on the basis of mathematical proofs and experimental evidence. Our belief in libertarian free will lacks any such impressive credentials. On the contrary, our enduring belief in indeterminist free will, like our belief in indeterminism for middle-sized objects, can easily be explained by our lack of discrimination.

This sort of genetic argument depends on certain philosophical assumptions about when we lack justification. But more importantly for present purposes, the argument depends on a speculative story about how we come to believe in libertarian free will. We need a well-confirmed psychological account to develop any such argument adequately. But I suspect that if we want to know in our lifetimes whether we should believe in libertarian free will, our best hope is a psychologically informed genetic argument.

THE PRESCRIPTIVE PROJECT

When we turn to the prescriptive project, the question is, "What should we do?" On this question, one might expect that psychology has absolutely nothing to contribute. After all, this is a question about *ethics.* However, I think psychology might make major contributions even here, for if knowing the facts will help us make better decisions about what we should do, the facts do make a contribution to prescriptive concerns. As noted in the first section, "Background: Three Projects," the prescriptive project is especially pressing for free will eliminativists. On that view, there is a fundamental mistake in lay notions of free will and responsibility. Two key questions that then emerge for the eliminativist are the following: (1) What *would* happen if people stopped believing in libertarian free will? and (2) Would a revolution be on balance beneficial? Psychology is poised to help with each question.

Some philosophers have worried that if people give up on free will, this might have dire effects on everyday life, and as a result, some have suggested that we ought to keep the truth hidden from the hoi polloi. This is actually a venerable strand of worry in ethics. A number of philosophers who embrace a utilitarian ethics maintain that there would be bad consequences if the man on the street actually knew that utilitarianism captured the truth about ethics. Hence, they counsel concealment. In his influential utilitarian treatise, Henry Sidgwick writes,

> …a Utilitarian may reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands. (1907, p. 490)

More colorfully, William Lycan writes, "I believe…firmly in some form of act-utilitarianism in ethics, but the sacred principle of utility itself forbids me even telling you this" (1987, p. 136, note 1). The worry is that people will behave badly if they come to believe utilitarianism. Hence, the utilitarian maintains that we should keep the truth secret.

Similarly, if we thought that anarchy and despair would ensue if people knew that there is no libertarian free will, this might count as a reason to resist informing the public of the truth. Such in fact is the view of Saul Smilansky. He writes, "Humanity is fortunately deceived on the free will issue, and this seems to be a condition of civilized morality and personal value" (2002, p. 500), and "there would be considerable room for worry if people became aware of the absence of libertarian free will" (2000, p. 505, note 7).

If people become convinced that they lack libertarian free will, would terrible consequences ensue? This is clearly an empirical question. And the rather limited empirical evidence suggests that we needn't fear a catastrophe. A recent experiment indicates that people's views about responsibility and determinism shift depending on whether determinism is presented as a remote possibility or a very live possibility. In the experiment, subjects read nontechnical descriptions of determinism. In one condition, subjects were told to imagine another universe that was deterministic, and in the other condition, subjects were told to imagine that *our* universe was deterministic. In the "other universe" condition, subjects tended to say that it's impossible for agents to be fully morally responsible in that universe; however, in the "our universe" condition, subjects tended to say that it would still be possible for agents to be fully morally responsible in our universe (Roskies & Nichols forthcoming). Thus, it seems that when people consider determinism as a genuine possibility for our world, they tend not to descend into anarchic moral views.

Research using a quite different methodology also suggests that if people became determinists, it would have little effect on their judgments of moral responsibility. Viney and colleagues found that college students who were identified as determinists were no less punitive than indeterminists (Viney et al. 1982) and no less likely to offer retributivist justifications for punishments (Viney et al. 1988). Although these results from Viney and colleagues are suggestive, the measure used for identifying determinists is flawed, and so here is another obvious place for further psychological research.

The final and deepest prescriptive question asks what we should do, all things considered, if we lack libertarian free will. Obviously, this question is so broad that I can scarcely do it any justice, but let's consider one social concern that looms large in the free will debate—retributive punishment. Retributivist approaches to punishment, which are central to our penal system, maintain that some people *deserve* to be punished because they are *guilty.* Several free will eliminativists have promoted a revolution in the penal system, suggesting that given the absence of libertarian free will, we ought to stop the practice of retributive punishment (e.g., Greene & Cohen 2004; Pereboom 2001). But it seems rash to try to overhaul these practices before we have some idea about the consequences of such a revolution. If Marxists had been more attentive to the psychological facts about human nature, we might have been spared some disastrous experiments in social engineering.

Once again, the natural place to look for insight is psychology. Retributive punishment comes naturally to us. This is nicely illustrated in a study by Haidt and Sabini (forthcoming) in which they showed subjects film clips that depicted injustices, and then subjects were asked to rate different possible endings. Subjects were not satisfied with endings in which the victim dealt well with the loss and forgave the transgressor. Rather, subjects were most satisfied when the perpetrator suffered in a way that paralleled the original injustice. That is, subjects preferred the ending in which the perpetrator *got what he deserved*.

Recent work in experimental economics indicates that these motives for retributive punishment might play an important positive role. Fehr and Gachter (2002) found that people will pay to punish those who defect in public goods games even when there is no reason to think that such punishment will produce benefits for themselves. This kind of punishment seems retributive in nature—as in Haidt and Sabini's study, the subjects here presumably think that the defector deserves to be punished (as opposed, say, to thinking that it will produce better utilitarian outcomes). Fehr and Fischbacher (2004) showed that punishment of this sort is extremely effective at establishing cooperation. In one such demonstration, participants play a public goods game in which they have the option to cooperate by contributing significantly to a common fund or defect by not contributing. In the first 10 trials, there is no opportunity to punish others, and contributions drop precipitously across these 10 trials. Then subjects are informed that they will be able to punish others in subsequent trials. Immediately after they receive this information, the level of cooperation takes a great leap, and after a few trials, the level of cooperation is about 90% of the maximum possible (Fehr & Gächter 2000). Thus, the presence of retributive punishment can secure cooperation, and merely knowing that others are in a position to exact retribution makes a person more likely to cooperate.

The foregoing represents only one line of research on punishment. But if the role of retributive punishment is as important as these studies suggest, then to promote the abolition of retribution seems a dangerous cause. It is especially important to recognize that retributive punishment is so effective partly because we are so receptive to it. Merely knowing about the availability of retributive punishment seems to improve people's behavior. Of course, we need to consider as well the potential costs of keeping retributive punishment. But it would be wise to wait for more evidence before fomenting a revolution.

CONCLUSION

In this chapter, I've sketched the complex intellectual geography of the free will debate. This geography includes three distinct dimensions. On the descriptive dimension, we want to discern the character of lay views concerning free will and the origin of those views. On the substantive dimension, we want to know how well the lay views capture the way things really are. And on the prescriptive dimension, we want to know how we should act in light of what we find out about the existence of free will. For each dimension, I've argued, psychology has an important role to play. Indeed, psychology holds great promise for advancing our understanding here. Philosophers have long tried to resolve the free will debate in isolation from psychological science, and despite the enormous ingenuity and effort that has been exerted, the purely philosophical investigations have not led to any kind of wide consensus. Psychology is poised to breathe new life into these issues. And one might hope that by bringing the resources of psychology to bear on the issue of free will, we can achieve the consensus that philosophy alone has failed to reach.

ACKNOWLEDGMENTS

This paper has benefited from helpful comments from John Doris, Josh Greene, Tamler Sommers, and Manuel Vargas.

NOTES

1. "Theory" is intended in a weak sense here. Roughly, any internally represented body of information will count as a theory in the intended sense.

2. Some maintain that compatibilism is true with respect to responsibility but not with respect to free will (see Fischer 2006).

3. Free will eliminativists embrace the strong claim that there is no such thing as free will. An alternative would be to maintain that although we are mistaken in our various beliefs about free choice, this doesn't mean that our concept of free will is empty (for discussion, see Vargas 2005). On such a view, one might say that just as we mistakenly thought that whales were fish, we also mistakenly thought that free will required indeterminism. In both cases we were wrong—whales aren't fish and free will isn't indeterministic—nonetheless there are whales and there is free will.

4. Even if this is right, it would be premature to conclude that people are consistently indeterminist about choice because different kinds of questions seem to provoke determinist responses (Nichols 2006a). Given the central goal of this chapter, I will set aside this important complication.

5. Indeed, on some views, we *never* perceive the causal powers that we presume to underwrite deterministic processes.

6. See also Woodward (1998).

7. Johnson's research program has confirmed and extended these results in several ways (see Johnson 2003).

8. The CHILDES database is a collection of transcripts of spontaneous language use by several children in family settings.

9. Of course if we had access to deterministic causal mechanisms underlying our decisions, we would not think them indeterminist. But the same goes for risks. When we have access to deterministic causal mechanisms, we don't think the outcome was indeterminist.

REFERENCES

Bargh, J. A., & Chartrand, T. L. 1999. The unbearable automaticity of being. *American Psychologist* 54, 462–79.

Bargh, J. A., Chen, M., & Burrows, L. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype priming on action. *Journal of Personality and Social Psychology* 71, 230–44.

Bargh, J., & Ferguson, M. 2000. Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin* 126, 925–45.

Baron-Cohen, S. 1995. *Mindblindness.* Cambridge, MA: MIT Press.

Bloom, L. 1973. *One Word at a Time: The Use of Single Word Utterances Before Syntax*. The Hague, Netherlands: Mouton.

Brown, R. 1973. *A First Language: The Early Stages.* Cambridge, MA: Harvard University Press.

Campbell, C. 1957. *On Selfhood and Godhood*. London: George Allen & Unwin.

Chartrand, T. L., & Bargh, J. A. 1996. Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology* 71, 464–78.

Clarke, R. 1993. Toward a credible agent-causal account of free will. *Nous* 27: 191–203.

Csibra, G. 2003. Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London B* 358, 447–58.

Double, R. 1991. *The Non-Reality of Free Will*. New York: Oxford University Press.

Fehr, E., & Fischbacher, U. 2004. Social norms and human cooperation. *Trends in Cognitive Sciences* 8, 185–90.

Fehr, E., & Gächter, S. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980–94.

Fehr, E., & Gächter, S. 2002. Altruistic punishment in humans. *Nature* 415, 137–40.

Fischer, J. 2006. *My Way*. New York: Oxford University Press.

Freud, S. 1961. *The Future of an Illusion*. Translated by J. Strachey. New York: Norton & Co. (Original work published 1927)

Gopnik, A., & Meltzoff, A. 1997. *Words, Thoughts and Theories*. Cambridge, MA: MIT Press.

Greene, J., & Cohen, J. 2004. For the law, neuroscience changes nothing and everything, *Philosophical Transactions of the Royal Society of London B* 359, 1775–85.

Haidt, J., & Sabini, J. forthcoming. What exactly makes revenge sweet?

Heider, F., & Simmel, M. 1944. An experimental study of apparent behavior. *American Journal of Psychology* 57, 243–59.

D'Holbach, P. 1970. *The System of Nature: Or, Laws of the Moral and Physical World*. Translated by H. D. Robinson. New York, B. Franklin. (Original work published 1770)

Hume, D. 1955. *An Enquiry Concerning Human Understanding*. L. Selby-Bigge (Ed.). Oxford: Clarendon Press. (Original work published 1743)

Johnson, S. 2003. Detecting agents. *Philosophical Transactions of the Royal Society of London B* 358, 549–59.

Johnson, S., Slaughter, V., & Carey, S. 1998. Whose gaze will infants follow? Features that elicit gaze-following in 12-month-olds. *Developmental Science* 1, 233–38.

Kane, R. 1996. *The Significance of Free Will*. New York: Oxford University Press.

Kant, I. 1956. *The Critique of Practical Reason*. Translated by L. Beck. Indianapolis: Bobbs-Merrill. (Original work published 1788)

Kuczaj, S., & Maratsos, M. 1975. What children *can* say before they *will*. *Merrill-Palmer Quarterly* 21, 87–111.

Kuhlmeier, V. A., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science 14*, 402–8.

Leslie, A. 1995. A theory of agency. In D. Sperber, D. Premack & A. Premack (Eds.), *Causal Cognition*. New York: Oxford University Press.

Lycan, W. 1987. *Consciousness*. Cambridge, MA: MIT Press.

MacWhinney, B., & Snow, C. 1990. The child language data exchange system: An update. *Journal of Child Language* 17, 457–72.

Machery, E., Mallon, R., Nichols, S., and Stich, S. 2004. "Semantics, Cross-Cultural Style." *Cognition*, 92, B1–B12.

Nagel, T. 1986. *The View From Nowhere.* Oxford: Oxford University Press.

Nahmias, E. forthcoming. Intuiting freedom: A response to Shaun Nichols. *Journal of Culture and Cognition.*

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research* 73, 28–53. Nichols, S. 2004a. The folk psychology of free will: Fits and starts. *Mind & Language* 19, 473–502.

Nichols, S. 2004b. *Sentimental Rules: On the Natural Foundations of Moral Judgment.* New York: Oxford University Press.

Nichols, S. 2006a. Folk intuitions about free will and responsibility. *Journal of Cognition and Culture* 6, 57–86.

Nichols, S. 2006b. "Imaginative Blocks and Impossibility: An Essay in Modal Psychology." In S. Nichols (ed.) *The Architecture of the Imagination.* Oxford: Oxford University Press, 237–55.

Nichols, S., & Knobe, J. 2007. Moral responsibility and determinism: Empirical investigations of folk intuitions. *Nous* 41, 663–85.

Nichols, S., & Stich, S. 2003. *Mindreading.* Oxford: Oxford University Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. 2001. Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review108*, 291–310.

Nisbett, R., & Wilson, T. 1977. Telling more than we can know. *Psychological Review* 84, 231–59.

Nucci, L. 2001. *Education in the Moral Domain.* Cambridge: Cambridge University Press.

O'Connor, T. 1995. Agent causation. In O'Connor (Ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will* (pp. 173–200). New York: Oxford University Press.

O'Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will.* New York: Oxford University Press.

Pereboom, D. 2001. *Living Without Free Will.* Cambridge: Cambridge University Press.

Premack, D. 1990. The infant's theory of self-propelled objects. *Cognition* 36 1–16.

Reid, T. 1969. *Essays on the Active Powers of the Human Mind.* Cambridge, Massachusetts: MIT Press. (Original work published 1788)

Roskies, A., & Nichols, S. forthcoming. Bringing responsibility down to earth.

Sachs, J. 1983. Talking about there and then: The emergence of displaced reference in parent-child discourse. In K. Nelson (Ed.), *Children's Language* (Vol. 4., pp. 1–18). Hillsdale, NJ: Erlbaum.

Scholl, B., & Tremoulet, P. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences* 4: 299–309.

Sidgwick, H. 1907. *The Methods of Ethics.* London: Macmillan.

Smilansky, S. 2002. Free will, fundamental dualism, and the centrality of illusion. In R. Kane (Ed.) *The Oxford Handbook of Free Will*. New York: Oxford University Press.

Sommers, T. 2005. *Beyond Freedom and Resentment: An Error Theory of Free Will and Moral Responsibility.* Ph.D. Dissertation, Duke University.

Strawson, G. 1986. *Freedom and Belief.* Oxford: Oxford University Press.

Strawson, G. 1994. The impossibility of moral responsibility. *Philosophical Studies,* 75, 5–24.

Vargas, M. 2005. The revisionist's guide to moral responsibility. *Philosophical Studies,* 125, 399–429.

Viney, W., Parker-Martin, P., & Dotten, S. D. H. 1988. Beliefs in free will and determinism and lack of relation to punishment rationale and magnitude. *Journal of General Psychology* 115, 15–23.

Viney, W., Waldman, D., & Barchilon, J. 1982. Attitudes toward punishment in relation to beliefs in free will and determinism. *Human Relations* 35, 939–49.

Wegner, D. 2002. *The Illusion of Conscious Will.* Cambridge, MA: MIT Press.

Wilson, T. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious.* Cambridge, MA: Harvard University Press.

Woodward, A. 1998. Infants selectively encode the goal object of an actor's reach. *Cognition* 69, 1–34.

3 Determined and Free

David G. Myers

Imagine two identical persons—individuals who are, in every imaginable way (genes, prior experience, current brain states, etc.) perfect clones of one another. If we now, in exactly the same manner, independently ask them a simple question—"Coffee or tea?"—will each necessarily respond the same ("Tea, please, with a spot of milk")? The question has but two possible answers: yes or no. As William James said, "The truth *must* lie with one side or the other, and its lying with one side makes the other false."[1]

An answer of "Yes, they must respond the same, because all possible influences are identical" assumes *determinism*. Human behavior may be too complex for you and me to predict the clones' responses. But their responses are, the determinist assumes, lawfully related to the interplay of causal influences. The two human copies, sitting in identical rooms, may each, after a moment's deliberation, have made a conscious choice that *felt* free to them. But an all-knowing being could have foreknown their decisions.

An answer of "No, humans are ultimately free" assumes at least some *indeterminism*, of inherent unpredictability. Much as elementary particles behave with seeming randomness, so human behavior will exhibit some intrinsic unpredictability. By this view, humans are, to at least a modest extent, free agents; they are a partial first cause of their own actions. After every conceivable biological, psychological, and social influence is accounted for, you and I retain the unpredictability and freedom to tip the scales this way or that, toward coffee or tea, and toward moral or immoral action. Your genes, your upbringing, and your circumstances may predispose certain behavior tendencies. But ultimately it is you who decides and who bears responsibility.

COMMON OBJECTIONS TO DETERMINISM

Would the implications of determinism be as troubling as some believe? Let's look at some common objections to it:

Does determinism cause fatalism? "The ideal reasoner," Arthur Conan Doyle's Sherlock Holmes once remarked, "would when he once has been shown a single fact in all its bearings, deduce from it not only all the chain of events which led up to it, but also all the results which would follow from it."[2] Sherlock expresses the seeming fatalism implicit in determinism. The universe is like a clock unwinding through its predictable course through time. Whatever will be will be, as the ideal reasoner could foresee.

Even so, determinism encourages us to action, not resignation. The stream to the future runs through our present acts. Our actions have effects; they help determine tomorrow's world. Because the past cannot determine the future apart from the present, this lays a responsibility upon us. Whether global warming becomes a weapon of mass destruction depends on our actions today.

If the opposite were true—if *in*determinism describes reality—then we *would* have cause for resignation. If today's events do not lead to tomorrow's world, then nothing you or I can do will make a predictable difference. So why bother? Responsible action assumes some predictability.

Does determinism deny free choice? We all experience choice. We weigh alternatives, make decisions, and, especially when we do well, accept responsibility. Indeed, among the animals, humans display enormously varied choices, from brutal genocide to self-sacrificial altruism.

Determinism does not for a nanosecond deny the practical freedom that people experience and cherish. Humanistic psychologist Rollo May could have been a determinist and still had his "free man"—a person who respects rational authority, acts responsibly, has self-respect, and is conscious of his or her part in decision making.[3] Determinism does not compel people to act against their will, nor does it deny them their experience of choice and their freedom to shape the future. The freedom guaranteed by democracies and for which people have died does not require indeterminism, but rather the right to make choices free of unreasonable restraint.

Does determinism negate praise or blame? My cognitive neuroscientist colleague Malcolm Jeeves recounts the case of a teacher who began collecting sex magazines and compulsively visiting child pornography websites.[4] When the teacher started making subtle advances to his stepdaughter, his wife called the police and he was arrested and convicted of child molestation. The day before his sentencing he went to a hospital emergency room with a distraught spirit and a severe headache. The medical staff found him "totally unable to control his impulses." He even propositioned the nurses. An MRI brain scan revealed an egg-sized tumor pressing on his right frontal lobe. When the tumor was removed, his lewd behavior ended. A year later, the tumor partially grew back and the man started once again to collect pornography. A further operation again removed the tumor and his urges again subsided.

In hindsight, we can surely agree that the teacher should not be held criminally responsible for the results of his brain tumor. But how far do we carry this line of reasoning? Should Mel Gibson, who in 2006 spewed anti-Semitic remarks to a police officer while under the influence of alcohol, be held responsible for his remarks? (Commentators wondered, was this alcohol talking, or was it Mel?) Should teens, whose immature frontal lobes exert limited control of impulses from their faster developing limbic system, be absolved of responsibility for slamming doors and unwise judgments? And what about those who, after living with abuse or with genes that predispose a reactive temperament, respond violently to provocation? Or those who, under more favorable circumstances, behave heroically? To pose the question generically, to what should we attribute someone's behavior—the conditions that bred it, or the person who enacts it?

But that is like asking whether we should credit the tree for the apples it bears, or the earth that bears the apples by bearing the tree. The earth may ultimately account for the apples, but that does not nullify the tree's accountability. If it bears few good apples, the farmer, who remains quite capable of praising what is good and condemning what is bad, will chop it down. And so with people. Our deeming their behavior blameworthy or praiseworthy does not require indeterminism.

Here, though, the wicket gets sticky. Although determinism leaves us free to judge *behavior* as worthy of praise or censure, does it leave us similarly free to judge the *person*? If a super hypnotist were to plant an irresistible suggestion of a crime in a person's mind which that person then committed with no sense of being coerced, would we hold the person responsible? To the extent that we are aware of an evil behavior's causes—whether a brain tumor, a hypnotic suggestion, or prior child abuse—we are less inclined to hold the person responsible. *Accountability presumes freedom,* argued Immanuel Kant.

This cuts the other way, too. We tend to credit and honor people for their benevolent acts only when we do not fully understand what led to them. (The Carnegie heroism awards are not given to those who rescue relatives.) If we had complete knowledge of the conditions that brought them to act as they did, we would shift credit to the conditions. Some years ago I heard B. F. Skinner remark that the defense of bad teaching is that it allows the student to be credited with learning. (With good teaching, he went on to explain, we're inclined to credit the teacher rather than the learner.)

The District of Columbia Court of Appeals expressed a judicial concern with the deterministic assumption that our judicial judgments of persons would be undermined if we assumed "that the behavior of every individual is dictated by forces—ultimately, his genes and lifelong environment—that are unconscious and beyond his control.... Our jurisprudence... while not oblivious to deterministic components, ultimately rests on a premise of freedom of will."[5]

So, human responsibility requires order and predictability, but also accountability for one's choices and actions. In what follows, I quickly review psychological science's explorations of (a) deterministic influences, and (b) the significance of human freedom, choice, and self-determination. I conclude by describing some parallel theological discussion of divine determination and human freedom, and by offering practical advice on when to assume determinism and when to assume freedom of the will.

PSYCHOLOGICAL SCIENCE CHALLENGES FREEDOM OF THE WILL

As psychological scientists, we operate from a working assumption that behavior arises from discernible causes. Within the complexity of human nature, there is order. When combined and interacting, our biology, our past experience, and our current situation powerfully influence our actions.

Evolutionary Influences

Evolutionary psychologists explore our shared human nature, especially the behaviors that allowed our distant ancestors to survive, reproduce, and send their genes into the future. We humans tend to fear snakes and spiders, and to avoid heights and bitter-tasting foods, because such fears and behaviors helped our ancestors survive. At the dawn of human history, our ancestors faced certain questions: Who is my ally, who my foe? What food should I eat? With whom should I mate? Some individuals answered those questions more successfully than others. As inheritors of their prehistoric genetic legacy, we are predisposed to behave in ways that promoted our ancestors' surviving and reproducing.

Genetic Influences

Behavior geneticists use twin, adoption, and sibling studies, and now microgenetic inquiry, to identify the contributions of genes and environment to our individual differences. With traits that range from general intelligence to temperament to extraversion, the predictability is often striking. In many respects, two identical twins, even if reared separately, are nearly as behaviorally identical as the same person tested twice.

Recent research is also specifying gene-environment interactions that predispose traits such as emotional disorder or antisocial conduct. One large study identified New Zealand young adults who had experienced several major stresses, such as a relationship breakup or a family death. These stressed individuals were much more likely to suffer depression *if* they carried a variation of a gene that codes for a protein that controls serotonin activity.[6] Although neither the gene nor the stress alone predisposed depression, the two in combination created the recipe for determining depression.

Brain Controls

Neuroscience studies reveal the tight bond of brain and mind. Everything psychological, it appears, is simultaneously biological. So, if mind is manifest by the physical brain, which operates by physical principles, can there be free will? Theologian Alan Torrance invites us to imagine that a complete description of the physical brain at Time 1, plus all incoming stimuli, could fully explain the brain state at Time 2. Would not freedom then become "a vacuous concept"?[7] Would not mind be but an epiphenomenon—a mere piggyback rider on the underlying brain events? Or, much as the material brain manifests immaterial consciousness, might free will be an emergent property of the brain's complexity?

Clearly, much human action arises apart from conscious free will. With selectively damaged brains, people may display *blindsight* (acting as if they see what they consciously cannot see). Given a split brain, a person may point to objects flashed to the nonverbal right hemisphere that are unknown to the verbal left hemisphere. It's as if there is another self within the body, whose experience is also evident in studies of *implicit memories* and *implicit attitudes* not available to the conscious self. That being so, where does freedom lie?

Neuroscientists are also identifying brain activity that foretells consciousness. In his provocative experiments, Benjamin Libet[8] observed that consciousness can lag behind the brain events that evoke it. When we move a wrist at will, we consciously experience the decision to move about 0.2 seconds before the actual movement. No surprise there. But our brain waves jump about 0.35 seconds ahead of our conscious perception of the decision. Thus, before we become aware of it, our brain appears headed toward our decision to move our wrist.[9] What may feel like a freely willed action has, it seems, been neurally preordained. (If Libet replayed the sequence in slow motion, he presumably could specify the exact time the participant is about to consciously will the movement.) The startling conclusion from these experiments is that consciousness (self-perceived freedom of the will) sometimes arrives late to the decision-making party.

Parental, Peer, and Cultural Influences

Freudian psychiatry and early behaviorism argued that parents determine their children's futures. (Recall John B. Watson's boasting, "Give me a dozen healthy infants…and I'll guarantee to take any one at random and train him to become any type of specialist—doctor, lawyer…beggar-man, or thief.") Parents do influence some areas of their children's lives, such as their manners and political and religious beliefs. But in other areas, such as personality, siblings' shared environment accounts for less than 10% of their differences.

Peers more greatly influence the development of many behaviors and attitudes. Let a child grow up in an environment in which their language is spoken with one accent by parents and another accent by their peers, and their adult accent will be that of their peers, not their parents.[10] And peers are just one facet of the wider culture, which includes media influences as well. The mark of our species—nature's great gift to us—is our ability to learn and adapt. We come equipped with a huge cerebral hard drive ready to receive many gigabytes of cultural software—all the behaviors, ideas, attitudes, values, and traditions shared by a group of people and transmitted from one generation to the next. If people living under the Taliban were to switch places for a few moments with their counterparts on a French Riviera beach, each would be instantly mindful of the determining power of culture.

Nonconscious Influences

As John Bargh documents in his chapter for this volume, one big challenge to our assuming that freely willed choices steer us through life comes from experiments showing that most of our everyday thinking, feeling, and acting operates outside conscious awareness, often "primed" by subtle influences. This big idea of contemporary psychological science "is a difficult one for people to accept," report Bargh and Tanya Chartrand.[11] Our consciousness is biased to think that its own intentions and deliberate choices rule our lives (understandably, because tip-of-the-iceberg consciousness is aware only of its visible self). But as Daniel Wegner's clever studies (this volume) compellingly demonstrate, consciousness often overrates (though sometimes underrates) its own control.

Take something as simple as speaking. Strings of words in unplanned sentences effortlessly spill out of our mouths with near-perfect syntax (amazing, given how many ways there are to mess up). We hardly have a clue as to how we do it. But there it is. Thanks to what Bargh and Chartrand call "the automaticity of being," we fly through life mostly on autopilot.

As I review such neuroscience, cognitive, and social science research for new editions of my introductory psychology text, and compare these insights with what William James could report in his then-masterful *Principles of Psychology*, I find the revelations sometimes breathtaking. Although human behavior remains, in the nitty gritty, substantially unpredictable, our understanding of the determinants of behavior has progressed dramatically. Compared to a century ago, we know ourselves much better.

PSYCHOLOGICAL SCIENCE AFFIRMS FREEDOM

"There are trivial truths and great truths," the physicist Niels Bohr reportedly said. "The opposite of a trivial truth is plainly false. The opposite of a great truth is also true." And so it is with human freedom, which psychological science challenges but also affirms.

Self-Determination and Personal Control

Studies confirm that systems of governing or managing people that enhance freedom will generally promote health and happiness. Some examples include the following:

- Prisoners given some control over their environment—by being able to move chairs, control TV sets, and operate the lights—experience less stress, exhibit fewer health problems, and commit less vandalism.[12]
- Workers given leeway in carrying out tasks and making decisions experience improved morale.[13]
- Institutionalized residents allowed choice in routine matters—what to eat for breakfast, when to go to a movie, whether to sleep late or get up early—may live longer and certainly are happier.[14]
- Homeless shelter residents who perceive little choice in when to eat and sleep, and little control over their privacy, are more likely to have a passive, helpless attitude regarding finding housing and work.[15]

The bottom line is this: We have, surmise Richard Ryan and Edward Deci, a deep-seated need for autonomy and self-determination.[16] When feeling free and self-determined, we generally flourish. People with a strong sense of personal control smoke less, wear seat belts more, make more money, more often practice birth control, resist conformity, and delay gratification. Moreover, when believing in our own freedom of the will, add Azim Shariff, Kathleen Vohs, and Jonathan Schooler (this volume), we also behave more morally and responsibly.

Additional studies of *intrinsic motivation,achievement motivation,perceived choice* in dissonance-related attitude change, *self-efficacy, learned helplessness,* and *reactance* (a motive to restore one's freedom when feeling coerced) further testify to the benefits of feeling free and believing in our own possibilities. These studies put recent psychological research squarely behind conceptions of human freedom, dignity, and self-control. The bottom line is this: *People benefit from experiencing freedom and from viewing themselves as free creatures.*

Although psychological research on the benefits of perceived freedom and self-determination is relatively new, the positive message is not. We find it in Norman Vincent Peale's 1950s best seller, *The Power of Positive Thinking:* "If you think in positive terms you will get positive results. That is the simple fact." Believe that things are beyond your control and they probably will be. Believe in your freedom to do something, and maybe, just maybe, you will.

Freedom Across Cultures

Does ever-increasing freedom breed ever-happier lives? Some freedom and control is better than none, says Barry Schwartz.[17] He notes that the "excess of freedom" in today's Western cultures contributes to decreasing life satisfaction, increased depression, and sometimes paralysis. Increased consumer choices have been a mixed blessing. After choosing among 30 brands of jam or chocolate, people express less satisfaction than those choosing among a half-dozen options.[18] This "tyranny of choice" brings information overload and a greater likelihood that we will feel regret over some of the unchosen options.

Some cultures more than others prize personal freedom. People in competitive, individualistic cultures have more personal freedom, take more pride in personal achievements, are less geographically bound to their families, and enjoy more privacy. Their less-unified cultures offer a smorgasbord of lifestyles and invite individuals to construct their own identities. These cultures also celebrate innovation and creativity, and they tend to respect individual human rights. Being more self-contained, individualists also feel relatively free to switch places of worship, leave one job for another, or even to leave their extended families and migrate to a new place. Marriage is often for as long as they both shall love.

Collective cultures embed people in strong social networks, with stronger expectations and less emphasis on personal choice. The individualized latté—"decaf, single shot, skinny, extra hot"—that feels so good to a North American in an espresso shop might sound more like a selfish demand in Seoul, note Heejung Kim and Hazel Markus.[19] Korean ads are less likely to emphasize personal choice, freedom, and uniqueness, and they more often feature people together.[20]

DETERMINISM, FREEDOM, AND RELIGION

The astute reader will have noticed that psychology's working assumption of determinism and its documentation of the benefits of empowerment and perceived freedom are compatible. Even if our actions were absolutely determined, we would nevertheless be free to choose consciously among alternatives. We could still know that our decisions matter and that society, for practical if not philosophical reasons, may hold us accountable. What determinism denies is not the practical consequences of our inner beliefs and choices, but the philosophical idea of agent causation—that people are *ultimately* self-determining.

Most people assume that religion requires a strong assumption of free will and its associated concepts of moral responsibility. Indeed, Christianity, Judaism, and Islam do emphasize human freedom and responsibility. "Choose this day whom you will serve," said the prophet Joshua to the people of ancient Israel.

Yet in emphasizing both determinism and freedom, psychology actually is closer to historic Christian theology than most people suppose.[21] Attacks on the idea that we are self-made people—that, thanks to our free will, we are independently capable of righteousness—have come not only from determinists but also from Augustine, Martin Luther, John Calvin, and Jonathan Edwards. Our conception of human responsibility, they argued, must not deny three attributes of God:

1. *God's foreknowledge.* Scripture portrays the selling of Joseph into slavery, the evil acts of the Pharaoh, Peter's denial of Jesus, Judas' betrayal, and the crucifixion as all the result of human choices that God anticipates. Such evidence moved Luther to conclude that "If we believe it to be true that God foreknows and foreordains all things; that He cannot be deceived or obstructed in His foreknowledge and predestination; and that nothing happens but at His will (which reason itself is compelled to grant); then, on reason's own testimony, there can be no 'free-will' in man, or angel, or in any creature."
2. *God's sovereignty.* Theologian Jonathan Edwards, who later served as Princeton University president, would not give so much as an inch to human free will, because to the extent that human will is indeterminant—spontaneous and free—God's plans become dependent on our decisions. But this, said Edwards, would necessitate God's "constantly changing his mind and intentions" in order to achieve his purposes. "They who thus plead for man's liberty, advance principles which destroy the freedom of God himself," for whom not even a sparrow falls to the ground apart

from his will. Nor are human will and God's will separate quantities that sum to 100%. Rather, agreed Augustine, "Our wills themselves are included in that order of causes which is certain to God." God works in and through our lives, our choices.

3. *God's grace.* Luther argued that the bondage of the will was an essential foundation for the doctrine of grace. By ourselves, he argued, we are unable to act righteously, to have faith, and to contribute to our own salvation. All credit belongs to God. Humility, not pride, is mandated. What then is left to free will? "Nothing! In truth, nothing!" insisted Luther. Calvin was just as forceful: Because the term *free will* "cannot be retained without great peril, it will…be a great boon for the church if it is abolished."

The divine determinism assumed by the doctrines of God's foreknowledge, sovereignty, and grace is *not* identical with naturalistic determinism. Yet Thomas Aquinas argued (in the words of Michael Novak) that "grace operates (except in the rarest cases) through the ordinary contingencies and processes of nature.…The whole environment, the whole 'schedule of contingencies' that constitutes history, is graced."

Whatever their differences, the scientific working assumption of determinism and the theological assumption of divine sovereignty converge in affirming human dependence on forces beyond our conscious knowledge, and without negating the significance of human freedom and self-determination.

PRACTICAL WISDOM: VIEWING SELF AS FREE AND RESPONSIBLE, OTHERS AS INFLUENCED

It is adaptive for us to view our own actions as substantially under our personal control. To view ourselves as passive billiard balls is a recipe for helplessness. Viewing ourselves as free and responsible agents is the groundwork for self-discipline and self-initiative. If exaggerating our freedom is conducive to pride, then negating our freedom is conducive to sloth. Thus it is important to recognize our capacity for self-determination and to help others see themselves as free agents, mindful that the stream of causation runs through our next choice. As Albert Bandura (this volume) reminds us, there is reciprocal determinism between persons and situations, much as brain-mind relations are reciprocal—occurring both bottom up and top down.

When we shift our perspective to others, it is productive to appreciate that their behavior is influenced by a myriad of factors. That assumption reflects the real power of biological constraints and of what we humans, with our fondness for "dispositional attributions" so often underestimate—the power of past and present situations. Appreciating such forces also provides the foundation for empathy and the hope of effective social reform. By contrast, belief in others' free will has been a predictor of prejudice.[22] If others are free agents, capable of rising above miserable circumstance and responsible for their own destiny, then we can surmise that impoverished people are responsible for their misery. With the assumption of total free will, it's a just world in which people get what they deserve (which also means one can accept personal credit for one's own achievements).

The adoption of differing assumptions for self and others can be seen in the Abrahamic faiths, which tend to view the self as free and the other as caused. When the holy texts address us directly, they emphasize our responsibility for our failings and our decisions. When talking to us about others, especially the poor and the outcast, it frequently advocates the complementary perspective: Act with compassion. "Judge not." Take the beam out of your own eye before worrying about the motes in others'.

The book of Proverbs, for example, admonishes self-control of one's passions, receptiveness to instruction, and hard work. But when it turns to our outlook on others, it admonishes concern "for the cause of the helpless." "Those who oppress the poor insult their Maker."[23]

Thus science and religion concur in affirming both a determined order and the benefits of perceived freedom. We are the products of our biological and social histories, and we are the architects of our futures.

NOTES

1. James, W. (1896). The dilemma of determinism, in W. James (Ed.) *The will to believe* (pp. 145–83). New York: Longmans, Green & Co.

2. Doyle, A. C. (1892). *The adventures of Sherlock Holmes.* New York and London: Harper & Brothers.

3. May, R. (1967) *Psychology and the human dilemma* (pp. 178–80). Princeton: Van Nostrand.

4. Jeeves, M. A. (2004). How free is free? Reflections on the neuropsychology of thought and action. *Science and Christian Belief, 16,* 101–22 (citing R. Swerdlow & J. Burns. (2002). Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign. *Archives of Neurology, 60,* 437–40).

5. American Psychological Association (1972). *Monitor, 3,* 4.

6. Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H. L., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 30,* 386–89.

7. Torrance, A. (2003). Developments in neuroscience and human freedom. *Science and Christian Belief, 16,* 123–37.

8. Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 12,* 181–87; Libet, B. (2004).

Mind time: The temporal factor in consciousness. Cambridge, MA: Harvard University Press.

9. Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

10. Harris, J. R. (1998). *The nurture assumption.* New York: Free Press.

11. Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being, *American Psychologist,54,* 462–79.

12. Ruback, R. B., Carr, T. S., & Hopper, C. H. (1986). Perceived control in prison: Its relation to reported crowding, stress, and symptoms. *Journal of Applied Social Psychology, 16,* 375–86.

13. Miller, K. I., & Monge, P. R. (1986). Participation, satisfaction, and productivity: A meta-analytic review. *Academy of Management Journal, 29,* 727–53.

14. Timko, C., & Moos, R. H. (1989). Choice, control, and adaptation among elderly residents of sheltered care settings. *Journal of Applied Social Psychology, 19,* 636–55; Langer, E. J., & Rodin, J. (1976). The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. *Journal of Personality and Social Psychology, 334,* 191–98.

15. Burn, S. M. (1992). Locus of control, attributions, and helplessness in the homeless. *Journal of Applied Social Psychology, 22,* 1161–74.

16. Ryan, R. M. & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality, 74,* 1557–85.

17. Schwartz, B. (2000). Self-determination: The tyranny of freedom. *American Psychologist, 55,* 79–88; Schwartz, B. (2004). *The paradox of choice: Why more is less.* New York: Ecco/HarperCollins.

18. Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79,* 995–1006.

19. Kim, H., & Markus, H. R. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology, 77,* 785–800.

20. Markus, H. (2001, October 7). Culture and the good life. Address to the Positive Psychology Summit conference, Washington, DC.

21. The theological material that follows, and sources for it, may be found in *Psychology through the eyes of faith,* 2nd edition, by David Myers and Malcolm Jeeves. San Francisco: HarperCollins.

22. Myers, D. G. (1978). *The human puzzle* (p. 256). New York: Harper & Row.

23. Proverbs 14:31, NRSV.

4 Self-Theories: The Construction of Free Will

Carol S. Dweck
Daniel C. Molden

The definition or nature of free will is ultimately a philosophical question; whether people *believe* they have free will is a psychological one; and whether people actually have free will is in the terrain somewhere in between. Psychologically speaking, free will is the perception of choice, agency, or self-determination. The more unconstrained people believe their actions are, the more free will they believe they have.

What factors determine the extent to which people see their actions as relatively constrained or unconstrained? Because much of people's perception of choice and agency comes from the way they mentally construct themselves and their world, it follows that different mental constructions will appear to place different constraints on people's actions. Thus, in the first part of our essay, we will show how people's *self-theories*—their conceptions of human qualities as fixed or as malleable—create different psychological worlds. These worlds place different constraints on their actions, leading people to different perceptions and experiences of free will. Interestingly, perceptions of free will within the different self-theories we describe mirror those of different philosophical traditions.

We will also show how the different self-theories lead people to different psychological solutions for issues allied with free will, such as issues of moral responsibility and blame. You will see that people's solutions are not always what the philosophers might predict. Philosophers have worried that the absence of free will would mean the absence of responsibility and therefore the inability to punish people for their harmful actions. However, we find that those people who are most committed to determinism (via the belief that human character is fixed and deterministic) are also most likely to believe in strict and direct punishment in the form of retributive justice (an eye for an eye).

Finally, we will turn to the issue of how much free will people may in fact have. Much of the philosophical debate on determinism and free will has turned on whether the laws of nature allow for free will. As Fischer and Ravizza (1998) express it, "…causal determinism is the view that all events can in principle be fully explained by reference to past states of the world and the laws of nature" or "causal determinism is the thesis that, for any given time, a complete statement of the facts about that time, together with a complete statement of the laws of nature, entails every truth as to what happens after that time" (p. 14). However, by the "laws of nature," philosophers usually mean the laws of physics, not the laws of *human* nature. Indeed, van Inwagen (1983) explicitly denies that psychological laws have a role to play in this issue.

To a psychologist, this seems surprising. Why should the question of whether people have choice, agency, or self-determination rest on how the physical world works rather than on how people work? It is interesting to note that a number of 18th-century philosophers were quite interested in the laws of human psychology as they bore on the issue of free will (and not only on the existence of free will but on the experience of free will as well; Harris, 2005). Although today many philosophical theories of free will involve the workings of motivation and self-regulation, philosophers do not look to the laws of human psychology for enlightenment.

We suggest that the laws of human nature (in the form of the nature and workings of human character or personality) may have an interesting role to play in the debate.[1] Specifically, we argue that the issue of free will may, at least in part, turn on questions of human personality and how best to conceive of it. If, for instance, human personality were fixed at birth or entirely molded by forces outside of our control, and if this personality determined people's actions, then there would be no free will (cf. van Inwagen, 1983). If, on the other hand, human personality were not only dynamic and malleable, but also were shaped at least in part by the individual, then free will in its deepest sense would exist (Ekstrom, 2000). The research on self-theories provides a perspective on the nature of human personality and how it works.

Moreover, as psychologists, we are interested not only in the yes–no answer to "Does free will exist?" but also, if the answer is yes, in the extent to which human beings may be able to exercise this power. Thus in the last part of the chapter, using philosophers' definitions of and criteria for free will, we use our perspective to reflect on the existence and extent of free will.

In summary, this chapter focuses first on the psychological worlds that shape people's experiences of free will and then on the implications of human personality and how it works for the exercise of free will.

BACKGROUND: HUMAN AGENCY AND THE EXPERIENCE OF FREE WILL

Side by side with philosophers' treatments of free will, there have been changes in social thought about human agency. Throughout much of history, most people viewed their fate as largely controlled by external circumstances—their station in life, epidemics, famines, climate, and the like. Then, starting in the Renaissance with the idea that human beings were perfectible, Western societies came to locate more and more power within the individual (Cassirer, Kristeller, & Randall, 1948; Heller, 1981). This revolutionary idea was spurred on by the Enlightenment and its notions of human rights, social improvement, and scientific advancement. According to the historian Peter Gay (1969), Bacon and Descartes both made dramatic breaks with historical fatalism. Bacon revived an old Roman saying, "Man is the architect of his fortune," and Descartes wrote of how science would make us "masters and possessors of nature." Rousseau, too, wrote repeatedly about our capacity for self-perfection (Gay, 1969). This tradition ultimately spawned an unprecedented belief in human agency.

Is this a good thing? Modern psychological research suggests that, at least within Western societies, belief in the power of the individual over the constraints of the environment predicts better psychological adjustment and greater personal success. As a rule, people appear to fare better with an internal versus external *locus of control* (Rotter, 1966), feelings of self-determination versus external constraints (Ryan & Deci, 2000), and use of *primary control* (direct, agentic action) over *secondary* control (adjusting to the environment or event without trying to affect it; Heckhausen & Shultz, 1995; Lazarus, 1991).

Yet our work has revealed that putting the locus of agency within the person, as opposed to the environment, can take different forms and does not necessarily lead to a greater sense of free will or personal agency (see Dweck, 1999; Molden & Dweck, 2006). This depends in large part upon the particular internal factors to which one accords causal power (see Weiner, 1985). In some cases, as we will show, it may simply be a case of trading belief in one form of determinism for another. Therefore, in considering the implications of people's self-theories for their beliefs about and experiences of free will, we must look beyond simple conceptions of internal "facilities" and external constraints (e.g., Hume, 1739/1960) and examine people's conceptions of what their internal facilities are made of.

SELF-THEORIES

We have investigated this issue by focusing on people's beliefs about whether basic personal attributes (such as intelligence or personality) are fixed and static traits or, instead, more dynamic qualities that can be cultivated (Dweck, 1999;

Dweck & Leggett, 1988). The former belief is termed an *entity theory*, because here the assumption is that human attributes are fixed entities that are not subject to personal development. The latter belief is termed an *incremental theory*, because here the assumption is that human attributes can be developed or changed incrementally through one's efforts.[2]

These theories are typically assessed with a questionnaire that measures people's agreement or disagreement with statements such as "People can do things differently, but the important parts of who they are can't really be changed" and "No matter what kind of person someone is, they can always change very much." Across many studies with diverse populations, we have repeatedly found that (a) most people readily endorse either an entity or incremental theory, (b) each theory occurs with roughly equal frequency, and (c) no one theory is consistently linked to people's ability level, education, or cognitive complexity (see Dweck, Chiu, & Hong, 1995). Furthermore, although individuals can strongly and stably hold one theory or the other (Robins & Pals, 2002), the theories can also be experimentally induced (see Dweck, 1999).[3]

As we will show, people's self- theories have a cascade of effects on their personal motivation, as well as on the ways they judge and treat others. Measuring people's stable self-theories has proven to be an effective means of predicting these effects. At the same time, experimentally inducing one theory or the other has shed light on the dynamic nature of belief activation and has established the causal role such theories play in creating these effects.

SELF-THEORIES AND BELIEFS ABOUT THE DETERMINISTIC INFLUENCE OF INTERNAL TRAITS

Research comparing entity and incremental theorists has uncovered marked differences, ones that have important implications for perceptions of free will (for reviews see Levy, Plaks, & Dweck, 1999; Molden & Dweck, 2006). Not only do entity theorists by definition believe in fixed traits, but they also believe that these traits directly cause behavior in a highly predictable way (Chiu, Hong, & Dweck, 1997; Hong, 1994). In contrast, not only do incremental theorists by definition believe in more dynamic, malleable traits, but they also believe that people's thoughts, feelings, and motivations—which they view as controllable—play the major role in causing their actions (Hong, 1994).[4] Thus, both theories give the major causal role to factors inside the person, but those internal factors for entity theorists are not amenable to personal control, whereas those internal factors for incremental theorists are far more susceptible to it.

A good illustration of the causal power entity theorists ascribe to traits comes from a study by Chiu, Hong, and Dweck (1997). In this study, entity and incremental participants were given information about the traits of two hypothetical people (e.g., Henry is more aggressive than Edward on average). They were then asked to predict how likely it was that Henry would act more aggressively than Edward in a different situation in the future. Entity theorists made strong predictions that people would consistently act in line with their traits, whereas incremental theorists' prediction of trait-consistent behavior was markedly and significantly lower (for related results, see Erdley & Dweck, 1993).

Research by Hong (1994) provides more insight into incremental theorists' causal beliefs. In one study, participants were asked to give explanations for a number of actions. Analysis of these explanations again provided evidence for entity theorists' focus on traits as the primary causes of behavior, but also revealed incremental theorists' focus on more dynamic psychological processes—beliefs, emotions, and motivations—as the primary causes of behavior (see also Chiu, 1994). Interestingly, research examining how entity and incremental theorists explain and predict the actions of *social groups* has produced highly similar findings (Levy & Dweck, 1999; Levy, Stroessner, & Dweck, 1998).

This differential focus on deterministic personality traits versus personal motivation is also evident in people's accounts of their own behavior. Research shows that when confronted with challenges or setbacks in their achievement of important goals, entity theorists again make sense of these outcomes in terms of enduring traits (i.e., a lack of ability), whereas incremental theorists make sense of these outcomes in terms of an insufficient display of desire or dedication (e.g., a lack of effort; Blackwell, Trzesniewski, & Dweck, 2007; Robins & Pals, 2002; see also Hong, Chiu, Dweck, Lin, & Wan, 1999).

Here, too, entity theorists expect trait-consistent behavior (poor performance) to continue into the future. Indeed, after setbacks, entity theorists have even been found to disparage displays of effort as futile and as further evidence that one lacks the inherent ability necessary for success (Blackwell et al., 2007). These types of causal judgments have been observed not only for failures on experimental tasks administered in the lab (Hong et al., 1999), but also for the genuine and prolonged challenges that people experience in the course of their lives (Blackwell et al., 2007; Robins & Pals, 2002).

Taken as a whole, these findings suggest that different beliefs about the nature of people's traits and abilities may profoundly alter people's potential for perceptions of free will in choices and actions. Entity theorists' greater emphasis on the deterministic influence of fixed internal traits could serve to give them a sense of a stable and predictable world, but at the same time, constrain perceived opportunities for choice and agency. In contrast, incremental theorists' greater focus on people's thoughts, feelings, and motivations as causes of action—factors they believe can be controlled—could serve to enhance perceived opportunities for self-determination. Philosophers have long discussed the existential implications of living in a world with more or less free will. Examining the psychological frameworks created by different self-theories may provide an opportunity to take a closer look into these worlds.

PSYCHOLOGICAL CONSEQUENCES OF SELF-THEORIES

Some philosophers have argued that people should relinquish the "illusion" of free will (e.g., Strawson, 1986). Although some contend that the loss of belief in free will would still leave plenty of room for people to pursue meaningful goals (Honderich, 1993; Pereboom, 2001), some have acknowledged that this loss would come at a psychological cost. Indeed, Smilansky (2002) insists that the illusion of free will must be maintained lest we lose "…our sense of achievement, worth and self-respect" (p. 482). Approaching this issue from the perspective of our research on self-theories, the question becomes: To what extent do entity theorists suffer from the constraints that their mindset places on self-determination?

Self-Determination Following Challenge and Failure

Many studies conducted across a wide variety of domains—intellectual achievement, relationships, sports, and business—suggest that entity theorists do indeed tend to suffer compared to incremental theorists (see Molden & Dweck, 2006; Dweck & Molden, 2005). This is most clearly illustrated in people's reactions to challenge and failure. Because, for entity theorists, failure signifies not only that their own traits and abilities are deficient but also that there is nothing they can do to alter or develop these traits, they tend to show *helpless* responses to such experiences (Dweck & Leggett, 1988). That is, they often report more negative emotional reactions, withdraw effort, and attempt to avoid similar tasks or situations in the future (Beer, 2002; Blackwell et al., 2007; Hong et al., 1999; Kammrath & Dweck, 2006; Knee, Patrick, & Lonsbary, 2003; Robins & Pals, 2002; Wood & Bandura, 1989). They may also show defensive self-protection rather than actively engage with a difficult task, even one that may be important for their future (Hong et al., 1999; Nussbaum & Dweck, in press; Rhodewalt, 1994).

In contrast, because, for incremental theorists, experiences of failure signify only that their effort was lacking or that they need to focus on developing the relevant attribute, they tend to show *mastery-oriented* responses (Dweck & Leggett, 1988). That is, they report less negative emotion and take active steps to rededicate themselves to their original pursuit (Blackwell et al., in press; Beer, 2002; Hong et al., 1999; Knee et al., 2003; Robins & Pals, 2002). These same helpless and mastery-oriented reactions have also been produced when people are directly taught an entity or incremental theory in an experimental situation, serving to underscore the causal link between the self-theories and reactions to setbacks (Kray & Haselhuhn, 2006; Martocchio, 1994; Wood & Bandura, 1989; Hong et al., 1999; see also Mueller & Dweck, 1998).

Thus, entity theorists tend to respond to difficulty by relinquishing agency, whereas incremental theorists tend to react by reasserting their agency. Do these different reactions make a difference for important life outcomes? In a longitudinal study by Blackwell et al. (2007), students' math achievement was monitored during their transition from relatively simple elementary school mathematics to more challenging junior high school mathematics. Although entity and incremental theorists did not differ in their math achievement when they entered junior high, incremental theorists soon began to earn higher grades than entity theorists and this disparity continued to increase over the next 2 years. The discrepancy in performance was found to result in large part from incremental theorists' belief in the efficacy of renewed effort and their choice to persist in the face of setbacks (see also Hong et al., 1999). Furthermore, several studies have now taught students an incremental theory and shown substantial increases in their motivation and grades or achievement test scores in the face of challenging curricula both in junior high school and in college (Aronson, Fried, & Good, 2002; Blackwell et al, 2007; Good, Aronson, & Inzlicht, 2003).

It is interesting to note that this pattern of helpless or mastery-oriented behavior has also been found with regard to people's attempts to exert influence over or change *others*. Heslin, Vandewalle, and Latham (2006) have shown that business managers who hold an incremental theory are more likely to provide extra coaching in an attempt to improve the performance of a struggling employee than are managers who hold an entity theory. Echoing these results, recent studies by Kammrath and Dweck (2006) have shown that incremental theorists are more likely to respond to conflicts with their romantic partners by taking active steps to influence the partner and solve the problem. In contrast, entity theorists, perceiving their partner to be immutable, are less likely to attempt a mutual solution and more likely to consider exiting the relationship when the conflict is serious (see also Chiu, Dweck, Tong, & Fu, 1997).

These findings indicate that entity theorists tend to show less active and agency-affirming responses to challenges and setbacks, but do entity theorists' more deterministic beliefs and less agentic behavior in fact result in the loss of self-worth that Smilansky (2002) fears? A longitudinal study on how people's feelings of self-worth change during their college years (Robins & Pals, 2002) has addressed this question directly. Results showed that entity theorists displayed a clear downward trajectory in their self-worth relative to incremental theorists. The diverging self-esteem of the two groups was directly linked to entity theorists' greater helpless reactions to failure, as compared to incremental theorists' active choices to persist.

Further support for these findings comes from recent research by Baer, Grant, and Dweck (2005).[5] They showed, first, that entity theorists experience greater symptoms of distress and depression in their daily lives and that this is tied to their greater tendency to engage in self-critical rumination about their fixed traits and abilities following negative events. Baer et al. also showed that the more distress entity theorists felt, the *less* they engaged in active problem solving. The opposite was true for incremental theorists. Overall, then, entity theorists' belief in deterministic traits leads them to perceive fewer choices for action following failure, and they do indeed appear to suffer from this lack of choice and reduced agency.

This is not to say that incremental theorists are immune to suffering and failure, or that belief in free will comes without its costs. It is possible that incremental theorists would be more vulnerable to dedicating themselves to lost causes, that is, situations in which withdrawal would be a better option than maintaining an "illusion of agency." Moreover, in circumstances where failures are not reversed by increased effort and dedication, incremental theorists may be even more dejected than entity theorists, who did not hold such hopes (see Beer & Srivastava, 2005; Plaks, Grant, & Dweck, 2005). Thus, whereas a belief in the absence of free will can have serious psychological consequences, one must also be wary about perceiving personal control where it does not exist.

SELF-THEORIES AND MORAL RESPONSIBILITY

The research described above provides strong indications that the more "deterministic" perspective of entity theorists can leave them feeling more helpless.[6] This raises a further question: Does the helplessness they feel after their own failures lead entity theorists to forgive other people's failures? That is, does their perception of a more fixed and deterministic personality compel entity theorists to absolve others of responsibility for harmful or misguided actions?

Although it can be readily argued that a decrease in responsibility follows from a diminished sense of free will (e.g., Strawson, 1986), there are those who propose that one can reject free will yet still believe in responsibility (or at least in holding others accountable for what they do; see Fischer & Ravizza, 1998; Honderich, 1993; Pereboom, 2001). Thus, even if people's unacceptable behaviors are seen to be a product of internal traits that they are powerless to change, they may still be condemned for these behaviors. Studies comparing judgments of responsibility made by entity and incremental theorists have intriguingly shown this somewhat counterintuitive pattern of results: Entity theorists tend to be *harsher* judges of people's transgressions than are incremental theorists.[7]

Research by Erdley and Dweck (1993) examined adolescent students' judgments of and desire to punish wrongdoers. After watching a slide show about a series of transgressions committed by a new boy in school, students who held an entity theory (a) showed greater moral condemnation of the boy and (b) recommended significantly more severe punishment than did incremental theorists (see Chiu, Dweck, et al., 1997, and Loeb & Dweck, 1994, for similar results with college students). Interestingly, entity and incremental theorists tend to rate a given transgression as equally wrong, serious, or negative (e.g., Chiu, Hong, et al., 1997), so it is not the case that incremental theorists simply have more lenient standards. It is when entity and incremental theorists come to judge the moral character of the transgressors and the punishment they deserve that they part ways.

Self-theories predict not just the amount of punishment people recommend, but also the type of punishment and its fundamental purpose. Loeb and Dweck (1994) had college students read scenarios and imagine themselves the victim of immoral or harmful acts. For entity theorists, retribution was the highest priority. They thought that the people who had harmed them should be harshly punished for the suffering and loss they caused. In contrast, for incremental theorists, education was the highest priority. Their hope was that the perpetrators could be educated about the harm they had done so that they might act more responsibly in the future (see Chiu, Dweck, et al., 1997 for similar results). In other words, whereas entity theorists focused on retribution, incremental theorists focused on reform.

This difference was further illustrated in research by Gervey, Chiu, Hong, and Dweck (1999). Participants, asked to play the role of jury members, were given a (fictitious) transcript of a murder trial. Entity theorists made stronger moral judgments of the defendant and used these character judgments more in their determinations of guilt, but what is most relevant for the present purpose is how the participants conceived of punishment and its purpose. Entity theorists believed that the most important function of the prison sentence was retribution for the wrongdoing, whereas incremental theorists believed that the most important function was rehabilitation.

In short, even though entity theorists believe that traits constrain the extent to which people could have acted otherwise, they still believe that people should be held accountable for these constrained actions. Moreover, even though incremental theorists believe that people are more free to alter and develop their basic character (i.e., they have *ultimate responsibility* for their actions; see Kane, 1996), they are less severe in their moral judgments and punishment. Perhaps it is their understanding of the complex psychology that lies behind decisions to act that makes them more understanding of errant behavior. Further, because in their eyes wrongdoers can change, they tend to advocate rehabilitation instead of retribution. To our knowledge, discussion of the link between conceptions of free will and the *type* of punishment people should receive has been less prominent in philosophical discussions and may be a fruitful direction for further analysis (see, for example, Smart, 1961).

To summarize thus far, research on self-theories has yielded a picture of two psychological worlds. In one, traits are fixed and deterministic and there is little room for agency when those traits prove deficient. It is also a world of retributive justice. In the other world, traits are malleable, and so are the causes of behavior, leaving more room for choice and agency even after setbacks. In this world of enhanced self-determination, education and rehabilitation are emphasized so that wrongdoers might be guided to make better choices in the future.

HOW DOES ENTITY-INCREMENTAL PSYCHOLOGY MAP ONTO THE PHILOSOPHY OF FREE WILL?

We now turn to the question of how entity and incremental theories might align with philosophical theories. Do people holding the different theories live in worlds described by different philosophers?

In the world of entity theorists, free will appears to be either absent or difficult and rare. This is because entity theorists (a) believe that human character and personality are fixed, (b) tend to view this fixed character or personality as the primary cause of action, and (c) do not tend to emphasize controllable motivational processes in the chain of influence. Thus entity theorists may have much in common with free will skeptics. Galen Strawson, a free will skeptic, argues that free will is impossible, in that it requires one to be a cause of oneself and nothing can be a cause of itself (Strawson, 1986). Paul Edwards (1958) says, "Ultimately our desires and our whole character are derived from our inherited equipment and the environmental influences to which we were subjected at the beginning of our lives. It is clear that we had no hand in shaping either of these" (reprinted in Kane, 2002, p. 62; see also Pereboom, 2001; Smilansky, 2000). Both Strawson and Edwards cite illustrious predecessors, such as Holbach, Schopenhauer, and Nietzsche[8] in support. Thus free will skeptics sometimes express a similar belief in a determined and determining personality to the one that entity theorists endorse.

Earlier, we showed that entity theorists, despite their deterministic leanings, were enthusiastic about retributive justice. As we noted, this may be seen as a contradiction. How can you punish people for actions over which they had no control? Some free will skeptics have wrestled with this problem and concluded that although free will is an illusion, it is a necessary one, one that we must not dispel. Without this illusion, they argue, we cannot hold people accountable for their actions and the foundations of society would be in jeopardy. As a case in point, Saul Smilansky (2000) begins his book *Free Will and Illusion* with the following statement: "This book explores the role of illusion in the free will problem, a problem at the heart of morality and human self-understanding.... Within our lives...[this] illusion is descriptively central and normatively necessary." (p. 1). Thus entity theorists (or some entity theorists) may be following Smilansky's recommendation. They do not believe in free will, but act as though people have free will and are morally responsible for their actions.

However, there is another possibility. A number of philosophers past and present have believed in a very strong degree of constraint on human behavior, but have not been comfortable thinking of people as deprived of their faculty for choice and agency. As a result, they have tried to create a system in which determinism and free will can logically coexist. It is possible that some entity theorists are like these *compatibilists,* that is, philosophers who believe that determinism does not rule out free will. Classic compatibilists, including Hobbes, Locke, Hume, and Mill, insisted that living in a determined world does not preclude the possibility of acting in ways we wish to act and leading the life we wish to lead. William James (1897, as cited in Kane, 2005) called this a "quagmire of evasion," suggesting that having it both ways is a bit suspect, but modern compatibilists seek to go even further. Not satisfied with mere freedom of *action,* they have tried to reconcile determinism with a deeper freedom of *will* (Kane, 2005). To do this, they have elaborated upon people's ability to make strong commitments to higher order desires (Frankfurt, 1971), to bring their desires into line with their reasons or values (Watson, 1988), and to understand right and wrong and act accordingly ("to do the right thing for the right reasons"; Wolf, 1988).

Much of the motivation of these compatibilists seems to derive from the fact that granting that at least some people have free will acknowledges that there are morally responsible beings. Frankfurt, Watson, and Wolf all reach for a kind of free will that makes people moral and responsible.

Entity theorists' simultaneous insistence on fixed, causal character and strict retributive justice could therefore also place them in the camp of the new compatibilists. Here people, determined though their characters and actions may be, still have the capacity and the imperative to act according to a set of moral standards—to obey higher order values, to bring desires into line with reasoned judgments, and to act in terms of right not wrong.

Finally, some entity theorists may even be libertarians, philosophers who believe in free will and view it as incompatible with determinism. Some libertarian philosophers portray acts of free will as extremely difficult and quite rare, requiring great exertion of effort or will at critical moments in the decision to act (e.g., Kane, 2005). For them, even if very few people ever perform such heroic acts of will, free will exists and has defeated determinism. In this view, self-regulation and moral behavior might be highly prized precisely because it is so exacting.

Where do incremental theorists' beliefs place them? Incremental theorists appear to fall in more easily with the free will affirmers because of incremental theorists' greater emphasis on (a) the contribution people make to the formation of their own character (b) the (in their view controllable) motivational processes that play a role in influencing action, and (c) the probabilistic rather than deterministic influence of personality on action. Thus incremental theorists could be allies of either compatibilists or libertarians. Indeed, incremental theorists seem to have much in common with those philosophers who argue strongly for the idea that people's character and values (not only their occasional actions) are at least in part determined by them and not by things beyond their control.

Specifically, *event-causal libertarians* argue that true free will and ultimate responsibility rest on *self-forming actions*—on our being the creators and originators of our purposes and ourselves (Ekstrom, 2000; Wiggins, 1998). Ekstrom (2000), for example, asserts that an agent is "constituted by a character, together with the power to fashion and refashion that character" (p. 113), and that this faculty for shaping character—for molding and harnessing our beliefs and desires—ought to seen as a major constituent of the self.

David Wiggins (1998), putting together the libertarian ideas of causal indeterminism and self-forming actions, proposes that what makes a human being different is not an unusual degree of causal power, but the fact "that his biography unfolds not only non-deterministically but also intelligibly; non-deterministically in that personality and character are never something complete, and need not be the deterministic origin of action; intelligibly in that each new action or episode constitutes a comprehensible phase in the unfolding of character, a further specification of what the man has by now become" (p. 293).

Thus incremental theorists and these philosophers share a belief in the power of human beings to perform self-forming actions, emphasize the chain of psychological-motivational events (beliefs, desires, intentions) that influence behavior, and subscribe to the probabilistic nature of the causal factors that produce behavior.

Perhaps incremental theorists would also be happy to align with Alfred Mele (1995, 2006), championing free will from both libertarian and compatibilist perspectives. Mele freely admits that libertarianism and compatibilism may seem (or even are) mutually exclusive, but he starts from the position that free will exists—that at least "some human beings sometimes act freely and are morally responsible for some of what they do" (Mele, 2006, p. 4)—and attempts to refute both antilibertarian and anticompatibilist arguments that would threaten its existence. So, in this interesting philosophical space, if either libertarianism or compatibilism is true, we have free will—although we will have a "better brand" of free will and moral responsibility if libertarianism is true. "Soft libertarians," as he calls himself and others like him, "would be disappointed to discover that determinism is true, but they would not conclude that no one has ever acted freely and that no one has ever been morally responsible for anything" (Mele, 2006, pp. 95–96).

Free will supporters and incremental theorists, by making people freer agents (with more control) in the chain of causes, give people moral responsibility. But because people also have self-forming power, it becomes a different kind of moral system, one with the emphasis on moral education rather than on retribution. That is, if people have the power to change their character and motivations, then, when they do something wrong, it makes sense to try to help them become better people instead of simply harming them in return. Indeed, Smart (1961), an *anti*libertarian, does find one (and perhaps only one) aspect of libertarianism attractive. Watson (2003) expresses it in the following way: "the idea of praise and blame as ways of *grading people with a view toward influencing their attitudes and conduct*, rather than *judging* them (where judging is understood as retributive in a punitive sense)" (p. 15).

In summary, if we examine the different philosophical treatments of free will, we find resonance with the worlds of entity and incremental theorists. Looked at from a different angle, entity and incremental theorists seem to be putting philosophical theories into action as they live their lives.

THE FIVE FREEDOMS: WHICH IS GRANTED BY EACH SELF-THEORY?

Another way to examine the way in which self-theories map onto philosophical theories of free will is through the lens of Kane's "five freedoms" (Kane, 2005). Kane points out that when different philosophers treat freedom of will they are not all talking about the same freedom. He then enumerates five meanings of freedom that emerge from different philosophical positions.

The first meaning is a simple "freedom of action." Championed by the classic compatibilists, it grants people the capacity to do what they wish to do. This view of freedom maintains that we are not prevented by external constraints from acting in ways that we desire to. However, according to Kane, this is a surface freedom, because it refers to freedom of action rather than a deeper freedom of will.

The second and third are freedoms emphasized by the new compatibilists and have to do with freedom of self-control. These freedoms involve people's ability to understand their own reasons and motives for acting (Freedom #2, là Frankfurt) or the right and good reasons for acting (Freedom #3, là Wolf) and to control their actions in accord with these reasons.

The fourth and fifth are the freedom of self-determination and self-formation, and involves the deepest sense of free will. This is the free will of the libertarians (and perhaps some compatibilists), in which people are acting on a will that derives from their character and motivations *that they played a role in creating*.

We propose that entity theorists who accept free will typically accept it in the sense of freedom of action or freedom of self-control. These freedoms are compatible with their more deterministic view and are sufficient to allow them to hold people accountable for their actions. They may also sometimes accept free will in the sense of self-determination if they focus on the idea that one can exert great effort and resist the strong forces of one's personality in order to act in a self-chosen way. Incremental theorists, on the other hand, can more easily subscribe to the fourth and fifth freedoms by virtue of their belief in deeper and more readily exercised forms of self-determination and in self-formation.

DO PEOPLE HAVE FREE WILL (AND HOW MUCH)?

Having considered the psychology of free will and the philosophical traditions it makes contact with, we now venture cautiously into new territory and ask, If it can be granted that the nature of personality has implication for free will, what are the implications of self-theories and the psychology of personality for free will? Thus, we enter the terrain between psychology and philosophy and use psychological findings to reflect on the question of free will. We venture there with the hope that our analysis may be interesting and thought-provoking.

In one sense, whether people have free will depends on which philosophical view and its criteria one adopts. Different theories specify different conditions for the existence of free will. For example, one can say with the compatibilists that even if personality and character were wholly fixed and deterministic, people could still exercise free will in the sense that their individual actions are uncoerced. Or one can say with the free will skeptics that even if personality and character were in some ways self-formed, nonetheless, at some point in the past (if only at the earliest phase of life) the individual did not have any say in laying the groundwork for the self. (Or that even self-formation springs from an impulse that is itself determined.)

Yet, in another sense, it matters deeply to our sense of free will whether personality is best thought of as relatively fixed and deterministic or relatively malleable (susceptible to self-forming actions) and probabilistic in its workings. But before we examine what our research findings have to say on this matter, it is useful to reflect upon what sorts of findings would and would not have implication for free will.

Most psychological research in the area of personality seeks to uncover the causes of our behavior. To the extent that this enterprise is successful, we will come to have a greater and greater understanding of the internal and external factors that make us who we are and that make us behave in lawful ways. Does this mean that the psychology of personality will inevitably pull the scales from our eyes and bring us face to face with our lack of free will? Kane (2005) cautions us not to confuse causation with constraints. Freedom of will is undermined by constraints—things that coerce us—not by causes. Free actions, he maintains, are unconstrained, not uncaused. This means that uncovering the causes of behavior need not endanger our sense of free will.

In a related vein, discovering predictability and lawfulness in human behavior does not imply determinism. We may measure certain personality factors and use our measures to predict people's behavior, but this does not mean that those factors were not, at least in part, self-formed or that those factors do not exert their influence in a probabilistic way. In other words, predictability and lawfulness do not rule out self-determination. That said, we will now use the findings from our research to reflect on the nature of personality and, in turn, its implications for free will.

The first point from our research is that personality is, in many ways, a highly dynamic system in which (changeable) beliefs can create a network of motivation and action (Cervone, 2004; Dweck & Legget, 1988; Molden & Dweck, 2006; cf.; Mischel & Shoda, 1995). For example, personal theories of intelligence create different goals, beliefs about effort, task choices, and reactions to setbacks. Although these beliefs can be instilled or activated without people's awareness, they can also be self-chosen.

We do not deny the important effects of inborn temperament and early experience and we do not view the child as a blank slate on which anything can be written. Yet our view of personality is very different from a view of personality as just a set of deep-seated personal qualities that inevitably incline people toward particular choices and actions. Indeed, even some of the most ardent students of temperament's role in personality grant that as children develop they cognitively construct their worlds, and that these cognitive constructions (such as self-theories) become an important part of their personality (Block, 1993; Rothbart & Ahadi, 1994).[9] Thus our first point is that people's belief systems are part of their personality, and we see in this more dynamic view of personality greater possibility for self-formation.

Our second point is that beliefs, such as self-theories, can alter what are often taken to be deep-seated traits. Such traits might include resilience, extroversion, openness to experience, conscientiousness, risk taking, and nurturance (Block, 1993; McCrae & Costa, 1999). Research has shown that self-theories, when adopted, can (a) modify these very personality traits and/or (b) modify the *effect* of such traits on people's choices and actions. The research that follows illustrates point (a).

- *Resilience:* Students taught an incremental theory displayed increased resilience in the face of real-world obstacles (Aronson, Fried, & Good, 2002; Blackwell et al., 2007; Good, Aronson, & Inzlicht, 2003).
- *Openness to experience and information:* When exposed to an incremental theory, managers were more open to information about changes in employee performance and were more open to negative feedback from employees that could help the managers improve (Heslin, Vandewalle, & Latham, 2006); students were more open to information about how to improve a disappointing performance (Nussbaum & Dweck, in press; see also Mangels et al., 2006); and people were more open to information that challenged their stereotypes (Plaks, Stroessner, Dweck, & Sherman, 2001).
- *Conscientiousness:* Students taught an incremental theory were more conscientious about doing their homework and studying for tests (Blackwell et al., 2007).

- *Risk taking:* Students exposed to an incremental theory were more willing to take on challenging tasks in areas of weakness than were students exposed to an entity theory (Nussbaum & Dweck, in press; Hong et al, 1999; see also Mueller & Dweck, 1998).
- *Nurturance:* Managers taught an incremental theory were more willing to give developmental coaching to employees, generated more coaching ideas, and generated higher quality coaching ideas (Heslin et al., 2006).

The evidence is more preliminary with respect to point (b), because in the relevant research, people's self-theories were not altered. Nevertheless, the theory people held about their shyness—independent of their *degree* of shyness—predicted the impact that their shyness would have on their decision to enter social situations and on their behavior once they were in the social situations (Beer, 2002). Those holding an incremental theory preferred more challenging social situations that would allow them to gain social skills, whereas those holding an entity theory preferred safer, less threatening situations. Those with an incremental theory were also more outgoing once in a social situation than those with an entity theory, resulting in more successful interactions. Thus the trait of *shyness* did not have a uniform and automatic effect on people's social choices and actions; rather, its impact was dependent on the self-theory people held.

To summarize this section, learnable beliefs are part of personality, they can influence other parts of personality, and they can change the relation between existing personality and action. If one accepts that the nature of human personality has implications for whether we have free will and can exercise it liberally, then this more dynamic view of personality may offer the opportunity for the easier and more frequent expression of free will.

CONCLUSION

In this chapter, our goal was to bring self-theories to bear on the issue of free will. We showed that people hold self-theories that have strong implications for their views on and experiences of free will and moral responsibility. We also showed that these self-theories play themselves out in people's actions in ways that suggest that a stronger belief in free will and agency can have benefits for people's achievement and well-being.

Further, we have demonstrated that the different self-theories with their different mental models of causality can be mapped in some interesting ways onto different philosophical theories of free will. Finally, we have argued that the self-theories provide a perspective on personality that may have implications for whether free will in an expansive sense, in fact, exists. Indeed, by illustrating the dynamic and potentially self-forming nature of personality, this perspective may lend weight to proponents of free will, with its emphasis on choice, agency, and self-determination.

In *Beyond Good and Evil*, Nietzsche (1886/1966) scoffed at the desire for free will "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" (p. 28). Few advocates of free will today would deny the importance of external powers or inherited influences. Many of today's philosophers who believe in free will ask only that *some* of our choices be freely made, that we have *some* say in who we are and what we do. Whether this wish can be granted—and how much free will can be granted—is not yet known, but it is to be hoped that psychological research can play a useful role in the debate.

ACKNOWLEDGMENTS

This chapter benefited greatly from the insightful comments of Manuel Vargas and we offer him our sincere thanks. Preparation of the chapter was supported in part by a grant from the National Science Foundation (Grant # BCS-02-17251) to the first author. Correspondence should be addressed to Carol S. Dweck, Department of Psychology, Jordan Hall, Stanford University, Stanford, CA 94305 or dweck@psych.stanford.edu.

NOTES

1. We will use the terms *character* and *personality* interchangeably.

2. We will use the term *self-theory* in this chapter to refer to beliefs about human attributes, not just one's own attributes.

3. The self-theories have been experimentally manipulated in several ways, for example, by giving participants reading passages that support a particular theory or by portraying the ability required by a task as fixed or acquirable.

4. Entity theorists tend to believe that the relevant thoughts, feelings, and motivations (as well as actions) issue from traits, and thus are not readily amenable to control.

5. Note that this research did not use a self-theories measure, but rather a closely related measure of people's emphasis on self-validation vs. self-growth. However, for clarity and continuity, we use the self-theory terminology.

6. Strictly speaking, one should not use the term *more deterministic*, because a system is either deterministic or it is not. We intend to convey the idea that entity theorists believe that they are subject to strong causal influences that make certain behaviors extremely likely.

7. As we discuss later, a belief in retribution need not be seen as incompatible with a belief in determinism (as in the view of some compatibilists) or with the

notion that free will may be very rare and difficult to exert (held by some restrictive libertarians). However, for the time being, we will pursue this line of thought.

8. Nietzsche is in some sense a free will skeptic as Strawson asserts, but he is also opposed to the thesis of determinism and the concept of an "unfree will," too, which he sees as resulting from the reification of cause and effect of natural science (see *Beyond Good and Evil*, section 21).

9. Temperament can, of course, affect the way in which people construct their worlds, but we have shown that changing one's self-theory nonetheless can exert a powerful influence on personality and behavior.

REFERENCES

Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–25.

Baer, A. R., Grant, H., & Dweck, C. S. (2005). *Personal goals, dysphoria, & coping strategies.* Unpublished manuscript, Columbia University.

Beer, J. S. (2002). Implicit self-theories of shyness. *Journal of Personality & Social Psychology, 83*, 1009–24.

Beer, J. S., & Srivastava, S. (2005). *Implicit self-theories, emotion, and social-approach.* Paper presented at the annual meeting of the Society for Personality and Social Psychology, New Orleans, LA.

Blackwell, L. S., Dweck, C. S., & Trzesniewski, K. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*. 246–63.

Block, J. (1993). Studying personality the long way. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & J. Block (Eds.), *Studying lives through time: Personality and development* (pp. 9–41). Washington, DC: American Psychological Association.

Cassirer, E., Kristeller, P. O., & Randall, J. H. (Eds.). (1948). *Renaissance philosophy.* Chicago: University of Chicago Press.

Cervone, D. (2004). The architecture of personality. *Psychological Review, 111*, 183–204.

Chiu, C. (1994). *Bases of categorization: The role of implicit theories in person cognition.* Unpublished doctoral dissertation, Columbia University.

Chiu, C. Dweck, C. S., Tong, J. Y., & Fu, J. H. (1997). Implicit theories and conceptions of morality. *Journal of Personality and Social Psychology, 73*, 923–40.

Chiu, C., Hong, Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology, 73*, 19–30.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality and development.* Philadelphia: Psychology Press.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*, 267–85.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256–73.

Dweck, C. S., & Molden, D. C. (2005). Self-theories: Their impact on competence motivation and acquisition. In A. Elliot & C. S. Dweck (Eds.) *Handbook of competence and motivation* (pp. 122–40). New York: Guilford Press.

Ekstrom, L. (2000). *Free will: A philosophical study.* Boulder, CO: Westview Press.

Erdley, C. S., & Dweck, C. S. (1993). Children's implicit theories as predictors of their social judgments. *Child Development, 64,* 863–78.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility.* Cambridge, UK: Cambridge University Press.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68,* 5–20. Reprinted in R. Kane (Ed.). (2002). *Free will.* Oxford: Blackwell.

Gay, P. (1969). *The enlightenment: The science of freedom.* New York: Norton.

Gervey, B. M., Chiu, C., Hong, Y., & Dweck, C. S. (1999). Differential use of person information in decisions about guilt versus innocence: The role of implicit theories. *Personality and Social Psychology Bulletin, 25,* 17–27.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An Intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24,* 645–62.

Harris, J. A. (2005). *Of liberty and necessity.* Oxford: Oxford University Press.

Heckhausen, J., & Schulz, R. (1995). A life-span theory of control. *Psychological Review, 102,* 284–304.

Heller, A. (1981). *Renaissance man.* New York: Schocken.

Heslin, P. A., Vandewalle, D., & Latham, G. (2005). The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology, 90,* 842–56.

Heslin, P. A., Vandewalle, D., & Latham, G. (2006). Keen to help? Managers' implicit person theories and their subsequent employee coaching. *Personnel Psychology,* 59, 871–902.

Honderich, T. (1993). *How free are you?* Oxford: Oxford University Press.

Hong, Y. (1994). *Predicting trait versus process inferences: The role of implicit theories.* Unpublished doctoral dissertation, Columbia University, New York.

Hong, Y. Y., Chiu, C. Y., Dweck, C. S., Lin, D., & Wan, W. (1999). Attributions and coping with challenges: The role of implicit theories. *Journal of Personality and Social Psychology, 77,* 588–99.

Hume, D. (1960). *A treatise of human nature.* Oxford: Oxford University Press. (Original work published 1739)

Kammrath, L., & Dweck, C. S. (2006). Voicing conflict: Preferred conflict strategies among incremental and entity theorists. *Personality and Social Psychology Bulletin, 32,* 1497–508.

Kane, R. (1996). *The significance of free will.* Oxford: Oxford University Press.

Kane, R. (Ed.). (2002). *Free will.* Oxford: Blackwell.

Kane, R. (2005). *A contemporary introduction to free will.* New York: Oxford University Press.

Knee, C. R., Patrick, H., & Lonsbary, C. (2003). Implicit theories of relationships: Orientations toward evaluation and cultivation. *Personality and Social Psychology Review, 7,* 41–55.

Kray, L. J., & Haselhuhn, M. (2006). *Implicit negotiation beliefs and performance: Experimental and longitudinal evidence.* Unpublished manuscript, University of California at Berkeley.

Lazarus, R. S. (1991). *Emotion and adaptation.* New York: Oxford University Press.

Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development, 70,* 1163–80.

Levy, S. R., Plaks, J., & Dweck, C. S. (1999). Modes of thought: Implicit theories and social understanding. In S. Chaikin, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 179–201). New York: Guilford Press.

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74,* 1421–36.

Loeb, I., & Dweck, C. S. (1994, June). *Beliefs about human nature as predictors of reactions to victimization.* Paper presented at the conference of the American Psychological Society, Washington, DC.

Mangels, J.A., Butterfield, B., Lamb, J., Good, C.D., & Dweck, C.S. (2006). Why do beliefs about intelligence influence learning success? A social-cognitive-neuroscience model. *Social, Cognitive, and Affective Neuroscience, 1,* 75–86.

Martocchio, J. J. (1994). Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. *Journal of Applied Psychology, 79,* 819–25.

McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–53). New York: Guilford Press.

Mele, A. (1995). *Autonomous agents: From self-control to autonomy.* New York: Oxford University Press.

Mele, A. R. (2006). *Free will and luck.* Oxford: Oxford University Press.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective systems theory of personality: Reconceptualizing the invariances in personality and the role of situations. *Psychological Review, 102,* 246–68.

Molden, D. C., & Dweck, C. S. (2006). Finding "meaning" in psychology: A lay theories approach to self-regulation, social perception, and social development. *American Psychologist,* 61, 192–203.

Mueller, C. M. & Dweck, C. S. (1998). Intelligence praise can undermine motivation and performance. *Journal of Personality and Social Psychology, 75,* 33–52.

Nietzsche, F. (1966). *Beyond good and evil* (W. Kaufman, Trans.). New York: Random House. (Original work published 1886)

Nussbaum, D., & Dweck, C. S. (in press). *Self-theories and modes of self-esteem maintenance.* Unpublished manuscript, Stanford University. *Personality and Social Psychology Bulletin.*

O'Connor, T. (2000). *Persons and causes.* New York: Oxford University Press.

Pereboom, D. (2001). *Living without free will.* Cambridge: Cambridge University Press.

Plaks, J.E, Grant, H., & Dweck, C.S. (2005). Violations of implicit theories and the sense of prediction and control: Implications for motivated person perception. *Journal of Personality and Social Psychology, 88,* 245–62.

Plaks, J., Stroessner, S., Dweck, C. S., & Sherman, J. (2001). Person theories and attention allocation: Preference for stereotypic vs. counterstereotypic information. *Journal of Personality and Social Psychology, 80*, 876–93.

Rhodewalt, F. (1994). Conceptions of ability, achievement goals, and individual differences in self-handicapping behavior: On the application of implicit theories. *Journal of Personality, 62*, 67–85.

Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1*, 313–36.

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology, 103*, 55–66.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 80 (Whole No. 609).

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Smart, J. J. C. (1961). Free will, praise and blame. *Mind, 70,* 291–306. Reprinted in Watson, G. (Ed.) (2003). *Free will* (2nd ed.). Oxford: Oxford University Press.

Smilansky, S. (2000). *Free will and illusion.* Oxford: Oxford University Press.

Smilansky, S. (2002). Free will, fundamental dualism, and the centrality of illusion. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 498–505). Oxford: Oxford University Press.

Strawson, G. (1986). *Freedom and belief.* Oxford: Oxford University Press.

van Inwagen, P. (1983). *An essay on free will.* Oxford: Oxford University Press.

Watson, G. (1988). Responsibility and the limits of evil: Variation on a Strawsonian theme. In F. Schoeman (Ed.), *Responsibility, character, and emotions.* Cambridge: Cambridge University Press.

Watson, G. (Ed.). (2003). *Free will* (2nd ed.). Oxford: Oxford University Press.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548–73.

Wiggins, D. (1998). *Needs, values, truth* (3rd ed.). Oxford: Oxford University Press.

Wolf, S. (1988). Sanity and the metaphysics of responsibility. In F. Schoeman (Ed.), *Responsibility, character, and emotions.* Cambridge: Cambridge University Press.

Wood, R., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of Personality and Social Psychology, 56*, 407–15.

5 Free Will, Consciousness, and Cultural Animals

Roy F. Baumeister

One afternoon, after you have completed your morning writing tasks and had your lunch, you head out to the car to do three errands in three different places. You need to take your dog to the veterinarian, pick up the specially ordered ice cream cake for your child's birthday party, and renew your automobile registration. Like most humans and animals, you would rather do the more enjoyable activities and put off the unpleasant ones, and in this case you will find doing something nice for your child to be pleasant, whereas separating from your beloved pet (who will gladly hop in the car but will turn recalcitrant upon recognizing the vet's office) will be unpleasant. So if you follow your gut impulses, you will start with the grocery store, then go to the DMV, and leave the vet till last. Wouldn't it be nice to have your sweet pooch welcome you back to the car after that tedious and annoying stint in the DMV?

But that plan has flaws, and if you think things through carefully, you might discard it. One flaw is that it leaves the dog and the ice cream cake alone together, unchaperoned, for the indeterminate amount of time you will be in the DMV, and so when you emerge with your new license plate, the dog will be too busy licking up the last of the ice cream to want to lick your face. And even if you make one change to avoid leaving the dog guarding the ice cream, there is another flaw, at least if you live in Florida as I do, which is that the temperature inside a car parked in the afternoon sun would seriously compromise the welfare of either a canine or any ice cream product. Hence the logical plan is to go to the vet first, then the DMV, and pick up the ice cream cake last so that you can spirit it directly home to the freezer with as little melting as possible.

This example is intended to illustrate one important concept of free will. In particular, you felt like doing things in one way but then overrode those impulses in favor of a different plan of action. Moreover, you employed conscious, logical reasoning, recognizing that the first plan would have unwanted, destructive consequences, and therefore worked out a second plan to produce a better outcome. Further, there is something distinctively human about this style of action (we know of no other animal who forms, evaluates, and revises plans in this manner), yet the benefits of that style of action are readily apparent, and so it may well be something distinctive about the evolution of the human psyche that created that capacity. And last, the formation of the second plan uses a style of reasoning that can be readily communicated to others, indeed discussed with them. If you told your spouse you were heading out to do the errands according to the first plan, he or she could point out the flaws, which most likely you would then recognize as a valid mandate for revision. Another person might offer you further information that would bring further changes, such as that the DMV happens to be closed today based on some obscure local holiday.

Free will is sometimes seen as a crucial, defining trait of human existence and other times regarded as an absurd, utterly implausible myth. Free will may be a vital faculty that human beings must exercise in order to fulfill their potential in creativity, virtue, or spiritual salvation. Or it may be a dangerous and obsolete illusion that all educated persons ought to reject immediately. The debate about whether free will is real is sometimes intense and bitter.

In this chapter, I will eschew the debate about whether free will exists. Instead, the focus will be on explaining the common belief in free will and, more important, on the phenomena to which those common beliefs refer. Rather than argue about whether free will is real, I shall focus on offering a model of how it might operate.

WHY NOT TO ARGUE ABOUT THE EXISTENCE OF FREE WILL

Philosophers, theologians, and others have discussed free will for centuries. Now psychologists have entered the picture. But perhaps arguing about the existence of free will should not be the job of psychologists. Instead, we might more usefully do what psychologists are most capable of doing, which is to test and refine theories about inner processes. Suppose there are two different kinds of processes that affect behavior, and one of them is freer than the other in some sense. Psychology might profit by exploring that difference, without getting bogged down in the highly abstract and philosophical debate about whether one of them satisfies the most rarefied criteria for qualifying as free will.

Though some experts balk at the term *free will,* few will dispute the view that human behavior can be guided by different kinds of processes. Moreover, concepts such as freedom, intention, and choice have not been rejected as obsolete, incoherent notions by psychologists. On the contrary, psychologists find those useful terms and produce significant differences in behavior as a function of them. Laypersons also recognize their own behaviors and others as sometimes voluntary and sometimes involuntary, they are acutely aware when their freedoms are threatened or removed, they judge themselves and others differently based on whether a behavior was freely chosen or coerced, and they struggle every day with making choices.

Perhaps the issue is moot, after all. Many philosophers embrace some form of compatibilism, which holds that one accept free will without relinquishing faith in causal determinism (e.g., Dennett, 2003; Kane, 2002). If many philosophers find the two beliefs compatible, psychologists should perhaps not fret that they will lose credibility as scientists if they, too, accept free will (cf. Wegner, 2002).

Personally, I resent being told that as a scientist I am forbidden to believe in free will and required to embrace total causal determinism. Possibly this resentment is caused by the well-documented motive to preserve the freedom to think what I want and entertain all possibilities (see Brehm, 1966). The reasons I offer for this refusal to be bullied, which admittedly may be different from my true or unconscious reasons (see below), are fourfold. First, the claim that all behaviors are fully determined by external, prior causes is unproven. Second, it is also unprovable, insofar as no one can study all possible behaviors and show them to be 100% caused by prior events. Third, it is contrary to our everyday experience. We all make choices every day, and it certainly feels as if more than one outcome is possible (hence the necessity of choosing)—which is precisely the point of free choice, namely that more than one could act differently. Fourth, it is contrary to our data, which almost invariably show probabilistic rather than deterministic causation. That is, behavioral outcomes in our experiments are almost never shown to be inevitable; rather, our causes reflect a mere shifting of the odds of some particular response.

In that sense, determinism requires a huge leap of faith, not unlike believing in a god. None of those objections proves free will to be true and determinism false. My point is simply that it is a form of unscientific fanaticism to require scientists to accept a belief that is unproven, unprovable, contrary to everyday experience, and contrary to their laboratory data.

Hence I propose to shelve the question of the existence of free will. Let us forge ahead and explore how action happens—and why people, perhaps rightly, perceive some actions as relatively freer than others.

WHAT NEEDS EXPLAINING

If the task is not to prove or disprove the existence of free will, then what? It is to understand how action is controlled and directed, and in particular to understand if there is some special model of action control that is (mostly) specific to humans and corresponds to the sort of action that people perceive, in themselves and others, as free.

Absolute freedom, in the sense that would meet the most stringent and exalted criteria for free will, is not necessarily the goal. Psychology mainly works by studying differences and degrees, and so if we can identify what differentiates the more versus less free actions, that would be a contribution appropriate to what psychology can do well. To be sure, if free will does exist in the most exalted philosophical or theological terms, it most likely would conform to what psychologists would identify as the relatively freer forms of action. In a sense, to map out the processes that produce relatively free action would be to furnish a model that would likely prove useful and relevant to understanding free will in whatever form turns out to be real.

The task can be further clarified by noting several likely parameters. In the first place, if there is no need to prove the existence of freedom, then certainly there is no need to depict all human action as free—if anything, the opposite assumption of partial, occasional, and incomplete freedom is more useful for psychology. The 20th century's most passionate advocate of free will was probably Jean-Paul Sartre (1943/1974), who famously argued that humans were "condemned to freedom" and who asserted that all human acts are free. In contrast, the approach I am advocating is to compare different (forgive the expression) *degrees* of freedom and to learn what distinguishes relatively free from relatively unfree acts. Under that assumption, free will is at best a sometime thing. Much, perhaps the majority, of human action could be fully and simply determined by simple, explicable causal processes, including brain dynamics, reinforcement-based learning, and ingrained or acquired responses to stimuli. Free will would represent only an occasional opportunity to suspend or override those causal processes, so as to allow a different process to take control.

The difference between seeing all human action as free versus seeing only occasional episodes of free action is rooted in different understandings of what is meant by choice, which can be roughly described as internal versus external. The external view of choice is defined by the situation: Different courses of action are technically possible. The inner view emphasizes the inner process. If the person does not go through an inner process of choosing, then the fact that in principle he or she could have done so is irrelevant. A person who always sleeps with the same partner or always eats the same food for breakfast is not making a choice every day according to the internal model, even though from an external standpoint one would say that those are choices insofar as the person could choose otherwise each day. At the extreme, the coffee shop franchise Starbucks has advertised that it offers 19,000 beverage options, but clearly this invokes the external criterion, because no one would want to make 19,000 decisions before having that first cup of coffee. For present purposes, we focus on the internal concept of choice. Free will is relevant only when there is an actual inner process.

That brings up the second point. The psychology of action has often struggled to understand how to make the transition from inner processes such as motivation and cognition to actual behavior: How, exactly, does thought cause the body to start to move? In this, it borrows the framing of the problem from physics, which has often sought to explain how a body at rest begins to move. But perhaps the problem has been wrongly phrased. Psychology, after all, has not identified any state of being "at rest." To the extent the free will exists, it serves not to initiate action so much as to alter and steer it. Put another way, behavior is already happening all the time, in all organisms from the simplest to the most complex. Free will does not therefore have to make behavior start happening, because there will be ample behavior without free will. Instead, the role of free will would be to alter the flow of behavior.

Viewed in that way, the function of free will is twofold. First, it overrides the response process that would happen without it. In other words, it suspends or interrupts one set of causal processes. Second, it chooses among various options for the coming course of action.

What sort of inner processes produce those outcomes? Almost certainly these include some degree of conscious, effortful thought. Insofar as behavior is purely produced by nonconscious, automatic processes, it would not be regarded as reflecting any sort of free will. From philosophical examples to laboratory studies, freedom of action is tied to conscious deliberation and intentional decision (Sartre, 1943/1974; Wegner, 2002), and that link is almost certainly maintained in everyday perceptions of freedom, such as legal decision making (i.e., if the person did not consciously intend to perform the forbidden act, then the responsibility and legal penalties are reduced). Hence some understanding of conscious processing is likely to be intimately linked to any ultimate theory of free will.

USEFUL FORMS OF FREE WILL

Many people seem to prefer to regard themselves as having free will. But why? Two decades ago, Dennett (1984) addressed the question of why free will would be worth having and suggested that some varieties would be more useful than others. In particular, he proposed that free will is relatively worthless unless it helps you get what you want. After all, having free will that does not make life better for you in any discernible manner would be somewhat pathetic, at best an idle form of random action devoid of practical benefit. Yet that definition may be something of a Trojan horse, because if people claim free will in actions that benefit them, the skeptics and determinists can almost always provide a causal explanation (e.g., you did that because your motives and wants caused you to do it).

Skeptics of free will can propose that they would accept an action as free if it were proven to have no relationship to any external cause or any prior event. To be sure, that sets the bar rather high, but it would at least constitute freedom in a sense that would be difficult to dispute. Then again, what sort of action would be utterly independent of external factors and prior events? In essence, it would be an essentially random action.

This view may also be responsible for some of the objections to the very idea of free will, especially by scientists. It treats free will as a random action generator. Science can accept randomness in nature, but deliberate and intentional randomness as a cause of human behavior seems absurd and postulating it seems a foolish and unproductive basis for any theory about behavior.

But would random actions help people get what they want? Or, to invoke another interpretive framework, would natural selection favor a capacity for random action? If we assume that evolution created human nature, then it is worth pausing to consider free will as a possible product of evolution (Baumeister, 2005; Dennett, 2002). If random mutations were to produce an increase in the capacity for purely random action, would those creatures survive and reproduce better than their competing, nonrandom peers? It is hard to see how natural selection would confer any substantial benefits on a random actor.

In contrast, natural selection might well confer huge advantages on creatures that developed other kinds of freedom. The preceding section contended that free will has a twofold nature, namely overriding a prepotent response tendency and making a smart choice among options. If a mutation offered a superior capacity for doing either of those, then the individual might well flourish (which means surviving and reproducing better than rivals), and the mutation would gradually spread through the gene pool. Let us consider each of those.

Overriding incipient responses is studied in psychology under the rubric of self-regulation (e.g., Baumeister, Heatherton, & Tice, 1994). Although there has been some work contending that certain self-regulation processes are nonconscious and effortless, for the present we are concerned with the conscious and effortful majority of self-regulation. Self-regulation is highly adaptive because it vastly increases the behavioral flexibility of the organism and hence its capacity to find an optimal response to a situation. Creatures that live in more complex and changeable environments, and who therefore encounter a broader diversity of situations, will benefit all the more from the capacity for self-regulation. Clearly, humans are at the extreme high end of that distribution (i.e., they live in very complex and changeable environments), and so a high capacity for self-regulation would be especially useful and adaptive for us. Sure enough, even just comparing humans against other humans, it is clear that people with a greater capacity for regulating themselves are more successful than their less self-regulating peers in a broad variety of outcomes (e.g., Duckworth &

Seligman, 2005; Kelly & Conley, 1987; Mischel, Shoda, & Peake, 1988; Shoda, Mischel, & Peake, 1990; Tangney, Baumeister, & Boone, 2004).

Self-regulation should qualify almost by definition as at least a limited form of free will. That is, without self-regulation, the organism cannot help but act on the first or strongest impulse that arises in response to a situation. With self-regulation, the organism can override that response, allowing a different impulse or response to take over. Overriding the first response frees the person from having to respond in that particular way and, if only briefly, creates a gap or uncertainty that opens the door for other possibilities. This is not to say that the eventual response is necessarily better than the first or that it is itself not the product of an inner causal sequence of responses. But the fact of changing away from the first to enable the second should constitute a kind of freedom, and it would almost certainly be recognized as such, though we are just now designing research to test that perception. And humans who could exert that much free will, who could override one response in order to permit another, would probably survive and reproduce better than their rivals who couldn't.

If self-regulation is one probably useful form of free will, rational choice would be another. Rational choice is an evolutionarily new kind of decision making (new with humans) that relies on evaluating a potential course of action using logical reasoning, most commonly with cost-benefit analyses that mentally simulate the various behavioral options and their likely consequences, quantifies them according to benefits to self-interest and possibly others, and compares them to produce the most desirable result (as far as can be ascertained with the available information). Undoubtedly it is facilitated by the use of language, which can represent the options, and by a mastery of logical reasoning, which enables the person to apply general principles to specific cases and to follow the ineluctable rules of logic to move from one thought to a different thought.

Rational choice is perhaps underappreciated in psychology, though some other social sciences (notably economics and political science) recognize it as a powerful model for understanding human behavior and if anything overstate how far it guides human behavior. Daniel Kahneman, who had a highly successful career based on showing how various behaviors and choices deviated from rational choice models, remarked in 2003 that he never intended to dispute the view of human beings as rational—people are rational, he said, just incompletely so. In my view, the incompleteness of human rationality confirms the depiction of free will as a sometime thing. The capacity for rational thought and decision making lies atop an irrational, impulsive beast, and so it only sometimes can alter the course of action that that impulsive beast will take.

The link between rationality and free will was noted by Searle (2001), who said that theories of rationality almost inevitably presuppose some degree of free will. After all, what would be the use of being able to reason out the best plan of action, if one were not capable of altering one's behavior so as to follow that best plan? Without that, the conscious mind would be a helpless, depressing spectator, constantly observing one's own irrational actions while figuring out that it would have been better to do something different.

The anecdote that opened this chapter illustrates the confluence of these two forms of free will (self-regulation and rational choice). One's preference and initial impulse dictate one sequence of action, but logical reasoning suggests that doing what one wanted would bring destructive consequences, and so one overrides the first plan in favor of a different and better plan. Humans make and follow such ad hoc plans constantly, but as far as we know that style of action control is absent in other species.

When philosophical writers such as Dennett (2003) discuss free will, their examples and discussions keep coming back to acts such as these, namely self-regulation and rational choice. It thus seems fair to regard them as two major manifestations of the sort of phenomena that are relevant to free will. Putting those two together gains plausibility, furthermore, from my own laboratory work. It appears that rational choice and self-regulation have overlapping inner processes that use a common resource.

The common link between self-regulation and rational choice emerged only gradually, in part because it pointed to a style of thinking that has been out of fashion for decades in psychology and only recently has begun to be taken seriously again. In this, I refer to theoretical models based on energy. Freud's theories characterized psychological processes as energy transactions, but more recent theorists have ceased to invoke energy, preferring highly cognitive theories and the like. The recent reconciliation between mind and body (driven by the rise of health psychology and neuroscience) has, however, furnished a newly plausible basis for using energy theories. The human body is, after all, an energy system that relentlessly ingests and burns calories. The special relevance of energy processes is suggested by Dunbar's (1998) observation that the human brain consumes 20% of the calories used by the entire body, while comprising only 2% of its total mass. Thus, the brain, and by implication psychological activity generally, is a huge burner of energy.

Our initial review of the research literature on self-regulation (Baumeister, Heatherton, & Tice, 1994) noted a pattern that suggested an energy process: It seemed that self-regulation operated as if depending on a limited resource. Self-regulation failure seemed more likely to occur when people had already self-regulated some other aspect of behavior. For example, when people struggle to quit smoking, they eat more, become crabby (failing to regulate emotion), and show other signs of poor self-regulation.

Although those observations were subject to rival interpretations, we began conducting experimental tests under rigorously controlled laboratory conditions, and we found that the pattern held up: After people engage in one act of self-regulation, they self-regulate less effectively in other spheres. For example, in one of the first studies, students who resisted the temptation to eat chocolate (and ate radishes instead, while staring at the chocolates) gave up faster on a subsequent discouraging task, as compared to students who had been permitted to eat chocolate or students who skipped the food procedure altogether (Baumeister, Bratslavsky, Muraven, & Tice, 1998).

The implication was that the initial act of self-regulation depletes some inner resource, leaving less available for the second task. This pattern has been widely replicated (for recent reviews, see Muraven & Baumeister, 2000; Schmeichel & Baumeister, 2004).

For present purposes, the most important finding is that logical reasoning in the service of making deliberate, conscious choices appears to deplete the same resource used for self-regulation. Thus, after making one big or multiple small choices, self-regulation is impaired (Baumeister et al., 1998; Vohs et al., 2006). Conversely, after people engage in self-regulation, their capacity for logical reasoning and decision making is impaired, as indicated by poorer performance on logic tests (Schmeichel, Vohs, & Baumeister, 2003) and increased reliance on fallible short cuts and heuristics in decision processes (Amir et al., 2005).

The common energy source is compatible with the view that evolution created a new or at least vastly renovated action control system for the human psyche. Essentially, evolution created a costly but powerful system by which the body's energy supply can be channeled into overriding behavior and making logical, rational choices.

Indeed, the evolutionary implications could be taken a notable step further. The fact that the psyche uses the same energy resource for self-regulation as for rational choice suggests that that mechanism evolved first for one of them, and then the second piggybacked onto the same system. That opens room for speculating about how the action control apparatus developed.

My best guess is that the system evolved originally for the sake of self-regulation, and it was later adopted and perhaps adapted for rational choice. The reasoning behind this is that self-regulation seems older and more fundamental in evolutionary terms than rational choice (especially logical reasoning). Long before humans evolved, social animals might need self-regulation in order to stifle their impulses and adjust their behavior to group life. For example, when a hungry animal sees food, the natural impulse would be to eat it. In a pack, however, if the animal begins to help itself to food before the alpha male eats his share, the smaller animal is likely to receive a beating. To continue to live in the group without daily beatings would require the capacity to overcome the natural impulse to eat those foods.

In plain terms, self-regulation allows the organism to alter its behavior so as to conform to the rules of the group. Logical reasoning and rational choice enable it to formulate its own rules, and the capacity for self-regulation can then be invoked to alter one's behavior accordingly. Both forms of free will promise to be highly adaptive.

We are beginning to explore the precise mechanisms behind this new, high-energy form of action control. Gailliot et al. (2006) found that blood glucose was reduced after acts of self-control and that giving laboratory a high-glucose snack counteracted the effects of ego depletion. Glucose is fuel for brain activities, though some (like self-control) use much more than others. The implication is that evolution developed a procedure for converting the human body's food energy into complex psychological processes that can alter the stream of behavior in adaptive ways.

HUMANS AS CULTURAL ANIMALS

Thus far I have suggested that free will, in whatever sense it actually exists, was the product of evolutionary processes that created the human psyche and that would have conferred advantages in survival and reproduction. An adequate psychological theory of free will should therefore situate it in the perspective of human evolution. The possibly quite special confluence of evolutionary pressures that produced the human psyche probably had much to do with the emergence of its novel system for action control.

In a recent book summarizing psychology's contribution to understanding human nature, Baumeister (2005) argued that the human psyche is distinctively well suited for participating in culture. The implication is that evolution took an unusual turn in creating humans. Social psychologists are fond of calling humans "the social animal" (see Aronson, 2000), but there are in fact many social animals. What sets humans apart is a radically new way of being social. Humans are most distinctively and appropriately described as cultural animals.

The sensory organs of most animals are geared toward detecting other species, mainly the wide assortment of predators and prey. In surprising contrast, human sense organs (especially those involved in vision and hearing) seem much more geared toward attending to each other. This is evident in the tradeoff between detection (noticing anything at all) versus resolution (processing a few things very clearly and thoroughly). Anyone who has lived with a dog knows it can hear a much wider range of sounds than a human—which would be useful for noting all manner of different animals—but the dog cannot distinguish similar sounds very well, so it cannot tell the difference between "Fido," "buy low," "hi ho," and so forth. The dog is not the unusual one here, though—it is the human ear, whose design has sacrificed the capacity for hearing high-pitched squeals and ultralow growls in favor of being able to hear subtle auditory differences between thousands of spoken words.

More generally, most animals get their food, shelter, and other needs directly from the natural environment around them. Humans get them from their social system. In fact, when faced with experimentally engineered conflicts between the evidence of their senses and the information given by other people, humans will often go along with the group rather than heed their senses (Asch, 1955, 1956).

In that sense, humans use a different biological strategy than almost any other species. Culture is the biological strategy that our species evolved to use, and most likely humans were selected on the basis of being capable of culture. Culture may be defined as an information-based system that enables people to live and work together in organized fashion to satisfy their biological and social needs. It offers huge advantages, which should be obvious given how thoroughly humans have taken over the planet and altered their environment. In simple terms of survival and reproduction, culture has been tremendously successful. In a relatively short time it has enabled humans to multiply from one woman to 6 billion souls, which is much greater success than any of our close biological relatives have had. In part this is because nearly all humanity's closest biological relatives live near the equator, but with the benefit of cultural learning to create clothing, cook food, and build complex shelters, humans have been able to live in a much wider range of physical environments than other species. And no other species has been able to triple its average life expectancy by means of its own research and interventions.

Culture has conferred advantages by several means, but all involve the social network (Baumeister, 2005). Language enables people to share information and make decisions in groups. Knowledge is stored in the social group rather than in the individual mind, so that people can benefit from the lessons learned and problems solved by people who are far away or even long dead, and this allows the collective body of knowledge to accumulate across generations, thereby creating progress of a sort that is almost entirely unknown in other species. Role differentiation and economic exchange enable social systems to become far more than the sum of their parts, including the fact that each task can be done by an expert, and trade can improve everyone's quality of life. (In fact, recent work has begun to suggest that Neanderthals failed to compete with *Homo sapiens* precisely because they failed to divide labor and hence were economically incompetent, which would indicate that evolutionary competition favored the humans who had the better cultural system; Horan, Bulte, & Shogren, 2005.)

In order to sustain and take advantage of culture, however, humans may have needed a new, more flexible and far-reaching psychological mechanism for making choices. Because of its potential for change and progress, humans encounter more different situations in their lifetimes, and a greater variety of choices, than other animals.

Culture can accumulate and use information most effectively by invoking meaning. Not coincidentally, humans therefore differ from other animals in the extent to which they use meaning to guide action. Ad hoc plans, such as the vet-DMV-supermarket errand example that opened this chapter, figure prominently in human activity but would be very difficult to develop and refine without language because they are based on integrating various ideas. The human brain evolved to become capable of language so as to be able to exploit the power of meaning. (The brain evolved; language was invented; meaning was discovered.) Again, though, such capabilities of thought would be largely useless unless people had sufficient free will to be able to alter their course of action based on those thoughts.

CONSCIOUSNESS AND SOCIAL LIFE

Earlier, this chapter suggested that if there are any genuine phenomena associated with the concept of free will, they most likely involve conscious choice. Such a view has to contend with the now widespread belief that consciousness is a useless, feckless epiphenomenon, and that all behavior is guided by nonconscious processes.

The attack on consciousness can be traced to Freud, who proposed that the conscious self is often merely and unwittingly carrying out the agenda laid down by unconscious motives and ideas. In recent years, the attack on consciousness has been spearheaded by John Bargh (e.g., 1994), whose careful and persuasive studies have shown that many ostensible results of conscious, intentional action can also be produced by activating nonconscious ideas or motives, thus entirely bypassing the conscious system. To be sure, those findings do not disprove the potential role of consciousness, and the argument that conscious will is dispensable fits well with my characterization of free will as a sometime thing, but it has been tempting for Bargh and others to speculate that consciousness will eventually be proven to be an irrelevance.

Further support comes from work by Gazzaniga (e.g., 2003), who has shown that the conscious mind devises explanations that are often fanciful and wrong. Likewise, Wegner (2002) has provided evidence that people are sometimes mistaken about whether they have caused something to happen, thus sometimes creating an "illusion of conscious will" (his term and the title of his book). Wilson (2002) has characterized conscious self-knowledge as likely to be useless and full of errors, and he has proposed that when people seek to explain their reasons for doing something, they are prone to fall into error and mislead themselves and others.

For many, the most devastating and influential critique of consciousness comes from the research by Libet (1985, 1999), who concluded that conscious thought is too slow to guide behavior. In his studies, participants were asked to initiate a simple motor action, moving a finger, and to note on a fast-moving clock when they made that decision. Their self-reports of the decision time were compared with electronic readings of brain activity, which showed that brain activity began to rise prior to the subjective decision time. In other words, the brain started to act before the conscious mind decided to act. To be sure, there are methodological critiques of that work (e.g., the subjective conscious time may reflect not the making of the decision but the self-recognition of having made the decision, which takes a bit longer), as well as arguments that is overstated and irrelevant (e.g., because consciousness compensates for its slowness by projecting slightly into the future; Shariff & Peterson, 2005). But the conscious processing system is known for being slow, and many researchers have come to accept the view that consciousness is too slow to guide behavior and is, hence, irrelevant to action.

In social psychology, the most influential critique of conscious thought in the modern era came from Nisbett and Wilson (1977), who contended that people cannot introspect on their thought processes (and are hence unable to explain their true reasons for acting). They said that when people are asked to explain their behavior, rather than introspect and furnish the actual, causal roots of their choices, they simply offer standard reasons from a stock of explanations that are favored in their culture. These explanations were derided as a priori theories about why people ought to make a particular choice, rather than true explanations of why they do.

Let us take these challenges seriously and offer a revised theory of conscious agency, based in part on the cultural animal perspective. First of all, from that view, humans evolved to work together with shared information (the basis for culture). Perhaps Nisbett and Wilson (1977) were overly disparaging of that stock of cultural explanations, treating them as trivial, irrelevant, self-deluding myths. Suppose that part of the value of consciousness is to enable people to make group decisions and to act in ways that would be suitable to the group. In that case, it is less important to come up with the true reason for one's action than to come up with a reason that the group will accept as justified. The local strongman may claim much of the best food because he is greedy and selfish, but he is likely to get less trouble if he can justify his claims to the group on the basis of the divine right and the gods' will. Today, a citizen defending his actions in a legal trial does not necessarily have to account for the true inner reasons for his actions but rather furnish an explanation for them that the law accepts.

In this view, then, consciousness is there in large part to help people explain and justify their actions, or to question and influence the actions of others, according to the collectively (socially and culturally) accepted rules. Well-designed studies can show that people sometimes are unaware of the subtle influences on their behavior, but the need to recognize those influences is less, or throughout our evolutionary history has generally been less, than the need to reconcile actions with the culture's rules.

Group decision making and interpersonal influence would thus benefit from consciousness, not because people necessarily know or acknowledge their true, inner reasons and causes, but because the group members can discuss and resolve the issue along mutually accepted lines. As a simple example, the true reason a parent desires good behavior from a toddler may be to prevent embarrassment to the parent, or to build character for decades hence, or simply to reduce the parent's hassle, but these reasons will be less persuasive to the toddler than the admonition that Santa Claus is watching and may reduce your allotment of toys for next Christmas. The culture sustains the useful fiction and the child believes it, and so the parent can invoke it to change the child's behavior. Thus, again, true reasons are less important than socially accepted ones.

Moreover, and perhaps even more profoundly, conscious thought can escape from the introspective ineptitude revealed by Nisbett and Wilson (1977), insofar as it masters logical reasoning and other rule-based forms of thought that are open to inspection. Perhaps, as in their famous study of stocking preferences, people were unable to introspect or even guess that they generally chose on the basis of a recency effect (picking the last stocking they saw) rather than on the basis of differences in color or texture. But the rules of logic are part of the objective reality of meaning, and intelligent people can agree with consensual certainty whether a conclusion follows from a premise. Thus, conscious thought introduced, or at least greatly promoted, a style of thinking that enables movement from one thought to another according to firm rules.

The errors, omissions, and blind spots with which researchers continue to lambaste consciousness may therefore be far less consequential than they first seem, because as people describe their conclusions to others, those errors and other flaws can be detected. Thus, as noted, Gazzaniga's split-brain patients may devise a false explanation for the ostensible coherence of different stimuli (the snow shovel and the chicken, in his standard example), but if they offered that explanation to others outside the laboratory, the others would quickly point out the error and correct them. Anyone who has tested children or even adults on arithmetic knows that not all conscious minds always reach the correct answer, but a correct answer does exist, and if a person can summarize his or her calculations to others, errors can be detected and corrected. The social network will thus correct the mistakes that the individual mind makes, as long as it can discuss them with other similarly conscious beings.

Likewise, in the example of the errand trip that opened this chapter, if you had indeed settled on the original plan and mentioned it to someone else, that other person would potentially have noted the dangers of leaving a dog and ice cream together in a car with an interior temperature over a hundred degrees.

Consciousness may therefore have developed as it did because it strengthened the link between the individual and the collective. It allowed the culture to guide the behavior of individuals in new and powerful ways. Because collective action through culture was the distinctive biological strategy of humans, anything that promoted it would likely help some humans prevail over their rivals in evolutionary competition.

USES OF CONSCIOUSNESS

The preceding section suggests that we should seek to understand the value and efficacy of consciousness in how it permitted a new form of action control that was suitable for cultural animals. That section emphasized finding reasons and justifications acceptable to the group and compatible with its rules. A related and potentially even more powerful reason has to do with logical reasoning.

As already noted, dual process theories of human mental functioning have now largely swept the field (e.g., Chaiken & Trope, 1999). Nearly everyone accepts that some cognitive processes involve automatic, nonconscious activity, whereas others are conscious and controlled. I have proposed here that free will, such as it is, will be mainly associated with the latter, although it may be more precise to speak of free will as arising from cooperative interplay between the two systems. Most conscious processing rests on a substantial amount of nonconscious activity. To furnish a simple and obvious example, understanding language requires a considerable amount of nonconscious processing, by which auditory or visual stimuli entering the brain from the sense organs are understood to convey meaningful ideas.

Earlier I suggested that rational choice based on logical reasoning was an important, adaptive, and common form of free will. How does logical reasoning occur? An influential article by Lieberman, Gaunt, Gilbert, and Trope (2003) proposed that it is largely confined to the conscious system (what they call "reflective"). To the extent that such reasoning occurs, therefore, it would be largely outside the capabilities of the automatic system, and consciousness can do it correctly.

A series of experiments in our laboratory has supported the hypothesis that effective logical reasoning depends partly on conscious processing (DeWall, Baumeister, & Masicampo, 2006). Several studies invoked the principle that the conscious system can generally do only one thing at a time (operates in serial), whereas the automatic system operates in parallel and therefore routinely performs multiple operations at once (Lieberman et al., 2003). Hence a distracting load would entirely preempt the conscious system but not the automatic system.

In several studies, we gave people logic problems to solve while listening to music. Some were instructed to monitor the music lyrics and count instances of a particular word; these showed substantial impairments in reasoning performance. In fact, they scored no better than chance guessing, though they seemed not to realize this and continued to answer questions as the same pace rather than simply putting down guesses rapidly for all items, a strategy that would have gotten them many more correct answers. Thus, preoccupying the conscious mind appears to have had a devastating effect on logical reasoning ability. Conversely, engaging the conscious mind more in the reasoning task, such as by telling people that they would have to explain their answers, led to significant improvements in scores on the logic test. Crucially, activating the idea of logical reasoning by means of nonconscious priming failed to have any significant effect on reasoning performance.

These findings offer preliminary support for the view that logical reasoning depends on a conscious processing system. Performance on the logic test rose and fell according to manipulations that targeted the conscious processing system, whereas manipulations aimed at the automatic processing system had no discernible effect.

Earlier I noted that Libet and others have emphasized the slow speed of conscious processing, as compared to automatic processes. Hence its operation may be to follow along and make careful corrections while the automatic process runs ahead and generates quick responses. This view is perhaps most compatible with Kahneman's (2003) characterization of "System 2" as a kind of editor that responds to the inputs from "System 1" (his term for the intuitive, automatic processes), sometimes accepting them, sometimes rejecting them or calling for revision. In one well-known study, Frederick (2005) asked people to tell the cost of a ball after hearing that someone had purchased a bat and a ball for $1.10 and that the bat cost a dollar more than the ball. Most people can get the correct answer of 5 cents, but usually their first thought is that the answer is 10 cents, and moreover people who are given the problem when distracted or in a hurry often give the 10-cent answer. The implication is that the automatic system processes the problem first and offers the approximate answer of 10 cents, and only the careful but slower operations of the conscious system comes up with the correct (and corrected) answer of 5 cents.

A similar view underlies our studies on creativity, which has periodically entered into the free will debate. Although some views of creativity see it as a form of free will, insofar as the person consciously decides how to formulate something new and different, the prevailing view among artists and psychologists has been that creativity is almost exclusively the product of nonconscious forces. The traditional emphasis on semidivine external muses as the wellspring of creative inspiration has largely given way to the assumption that creativity springs from deeply unconscious roots and that the conscious mind is, if anything, an impediment or obstacle to the creative process (see Dennett, 2003, and Wegner, 2002, for summary of some of those views).

But then why is creativity mainly found among conscious beings? We reasoned that perhaps the creative inspirations do emerge from nonconscious processes, but the integrative editing of the conscious mental system is vital for fashioning the final product. Hence preoccupying or distracting the conscious mind would reduce creativity rather than facilitating it.

A series of laboratory studies supported that hypothesis that creativity depends in part on conscious activity (Baumeister, Schmeichel, DeWall, & Vohs, 2007). In one, we asked musicians to perform a series of improvisations, one of which was done while counting backward by 6 from 917 (a cognitive load manipulation designed to preoccupy the conscious system). Judges rated those solos as less creative than solos done while counting forward by 1 or while not counting at all. In other studies, participants drew pictures while listening to music as in the logic studies above, either tallying instances of a lyric or just listening, and judges rated the drawings made during cognitive load as less creative than the others. It is important that not all aspects of performance were impaired by the cognitive load. When the conscious system was distracted, musicians were able to keep the beat and avoid mistakes (defined as notes outside the key), and artists successfully followed the instructions about what elements to include in their drawings and used the same variety of colors. But such aspects probably can be achieved automatically and hence do not rely on consciousness. Only the artistic integration into an appealing, creative product suffered when the conscious mind was otherwise engaged.

The special value of conscious functioning seems thus to be found in editing mental operations: criticizing them, combining them, and the like. The conscious mind may thus react to the impulses arising from automatic, nonconscious processing. Libet (1985, 1999) himself did not conclude that consciousness is irrelevant, only that its main function was to exercise a veto over behaviors that the automatic system initiated. Self-regulation is likewise most commonly exercised to stop a behavioral response sequence, rather than to initiate one (Baumeister, Heatherton, & Tice, 1994). But does that mean free will is merely reactive ("free won't" in the phrase of some writers; Ohbi & Haggard, 2004)?

THE DELAYED EXECUTOR

Let us return to the problem, raised by the Libet findings, that consciousness seems too slow to guide behavior. By the time the conscious system makes up its mind what to do, the ship has already sailed. Is there any way that such a slow conscious system could participate in guiding action?

Assume for the sake of argument that the immediate control of behavior is always the result of nonconscious processes. In the heat of the moment, the person's automatic system sizes up the stimulus situation and organizes the response. To do that, it presumably consults some bank of inner programs (including reinforcement history and other knowledge) that tell it how to respond to that situation. Nothing in that process requires participation by the conscious system.

Crucially, however, the conscious system could still have an indirect—yet extremely powerful and adaptive—influence over behavior if it can alter those programs. And for this, its slowness may be irrelevant, because the creation or revision of programs could be done in relatively quiet moments far removed from the crucial seconds when responses are being made. In particular, conscious reprogramming could occur after major events have ended (thus mainly helping to revise how one would respond to similar events in the future), or well before a particular, anticipated situation is encountered.

Unlike robots and computers, which go on to the new task as soon as the current one is finished, the human mind has a tendency to dwell on prior events and ruminate about them. In particular, unpleasant conscious emotions (which are usually a sign that some episode has turned out badly) stimulate the mind to ruminate about how the event might have gone differently, a pattern called counterfactual thinking (Roese, 1997). Such thinking seems ideally suited to this reprogramming function, which can be called the *delayed executor,* because the conscious mind examines the episode step by step to consider how a more desirable outcome might have ensued if one had responded differently at some point. Thus, as the event was unfolding, the conscious mind was perhaps not involved in directly steering the action, but it may have been gathering observations to use in its postmortem analysis. The automatic system dictated the responses according to its programming, and afterward the conscious mind weighed the outcome and mentally simulated different actions that might have been taken, and if it concludes that a different act would have yielded a better result, it essentially revises the program that the automatic system will consult in future similar episodes. Next time, take an umbrella, or get the promise in writing, or post your name inside the suitcase, or refrain from making commitments while inebriated.

Likewise, the organization of behavior according to ad hoc plans can be done by the conscious mind well in advance of their execution, so that the automatic system is fully in charge on carrying out the behavior. To return to the example that opened this chapter, the plan for the sequence of errands was made by means of conscious deliberation, before the errands were begun. Hence any researcher who studied the person carrying out those errands might find, correctly, that each act in the process of doing the errands was directed by automatic, even nonconscious responses, because at each moment your response was a direct, preprogrammed response. But the macro program had been crucially shaped by conscious thinking.

The relevance of conscious volition to macro thinking has been argued persuasively by Donald (2002), who argued that the cognitive science approach of studying ever more micro units of behavior will bias the data against finding any role for consciousness. At the extreme, by studying behavior at the level of neurons firing, one could explain behavior with no possible role at all for conscious processes, and researchers who work at that level might easily convince themselves that they had fully resolved the debate by ruling out any causal role for consciousness. But if one looks at behavior in larger units over longer periods of time, consciousness might just turn out to be decisive.

CONCLUSION

Most scientists reject the idea of free will as a random action generator, and probably with good reason. In contrast, free will in the sense of self-control and rational, intelligent choice comprises an important set of psychological phenomena and is plausible in terms of the evolution and construction of the human psyche. Quite likely human conscious processing emerged as a way to facilitate this new form of action control. It may operate less by direct initiation of behavior than by macro and sometimes delayed reflecting on optimal courses of action, possibly setting up and altering response tendencies that guide the automatic responses that are the immediate, proximal causes of behavior.

REFERENCES

Amir, O., Dhar, R., Pocheptsaya, A., & Baumeister, R. F. (2005). The fatigued decision maker: Ego depletion changes decision process and outcome. Unpublished findings, Yale and Florida State Universities.

Aronson, E. (2000). *Social animal* (7th ed.). New York: Freeman.

Asch, S. E. (1955, November). Opinions and social pressure. *Scientific American*, 31–35.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs, 70*, (No. 416).

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Baumeister, R. F. (2005). *The cultural animal: Human nature, meaning, and social life.* New York: Oxford University Press.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*, 1252–65.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation.* San Diego, CA: Academic Press.

Baumeister, R. F., Schmeichel, B. J., DeWall, C. N., & Vohs, K. D. (2006). Is the conscious self a help, a hindrance, or an irrelevance to the creative process? *Advances in psychology research*, *53*, 137–52.

Brehm, J. (1966). *A theory of psychological reactance.* New York: Academic Press.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology.* New York: Guilford.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

Dennett, D. C. (2003). *Freedom evolves.* New York: Viking/Penguin.

DeWall, C. N., Baumeister, R. F., & Masicampo, E. J. (2006). Evidence that logical reasoning depends on conscious processing. Manuscript submitted for publication, Florida State University.

Donald, M. (2002). *A mind so rare: The evolution of human consciousness*. New York: Norton.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*, 939–44.

Dunbar, R. I. M. (1998). The social rain hypothesis. *Evolutionary Anthropology, 6*, 178–90.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.

Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., & Plant, E. A. (2006). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. Manuscript in preparation, Florida State University.

Gazzaniga, M. S. (2003, January). The when, where, what, and why of conscious experience. Presented to the National Institute on the Teaching of Psychology, St. Petersburg Beach, FL.

Horan, R. D., Bulte, E., & Shogren, J. F. (2005). How trade saved humanity from biological exclusion: An economic theory of Neanderthal extinction. *Journal of Economic Behavior and Organization, 58*, 1–29.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kane, R. (2002). *Oxford handbook of free will.* New York: Oxford University Press.

Kelly, E. L., & Conley, J. J. (1987). Personality and compatibility: A prospective analysis of marital stability and marital satisfaction. *Journal of Personality and Social Psychology, 52*, 27–40.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavior and Brain Sciences, 8*, 529–66.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies, 6*, 47–57.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology, 34*, 199–249.

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*, 687–96.

Muraven, M. R., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*, 247–59.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–59.

Obhi, S.S., & Haggard, P. (2004). Free will and free won't. *American Scientist, 92*, 358–65.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133–48.

Sartre, J.-P. (1974). *Being and nothingness.* Secaucus, NJ: Citadel. (Original work published 1943)

Schmeichel, B. J., & Baumeister, R.F. (2004). Self-regulatory strength. In R. Baumeister & K. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 84–98). New York: Guilford.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Ego depletion and intelligent performance: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology, 85*, 33–46.

Searle, J. R. (2001). *Rationality in action.* Cambridge, MA: MIT Press.

Shariff, A., & Peterson, J. B. (2005). Anticipatory consciousness, Libet's veto, and a close-enough theory of free will. In R. Ellis & N. Newton (Eds.), *Consciousness & emotion* (pp. 197–215). Amsterdam, Netherlands: J. Benjamins.

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology, 26*, 978–86.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*, 271–322.

Vohs, K. D., Baumeister, R. F., Twenge, J. M., Schmeichel, B. J., & Tice, D. M. (2006). Decision fatigue exhausts self-regulatory resources. Manuscript submitted for publication, University of Minnesota.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Harvard University Press.

6 Reconstrual of "Free Will" From the Agentic Perspective of Social Cognitive Theory

Albert Bandura

The doctrine of "free will" was promoted by medieval theologians to explain the prevalence of evil given a benevolent and omniscient Creator, and has been debated by philosophers for millennia. Proponents of free will granted humans the power of free choice in the likeness of absolute agency. Free will is an enigmatic, autonomous causative force shrouded in conceptual ambiguity about what it is, where it comes from, and how it manages to operate autonomously in the midst of environmental pressures. The metaphysical analytic preoccupation with the incompatibility of free will and determinism diverted attention from more fruitful analysis of the capacity of humans to bring their influence to bear on events (Nahmias, 2002). Reframing the issue of free will in terms of the exercise of agency, operating principally through cognitive and other self-regulatory processes, holds greater promise of providing new insights into the constructive and proactive role that cognition plays in human action.

The capacity for personal influence must be analyzed in the broader context of the model of human nature in which it is rooted. The conceptions of human nature regarding the capacity to exercise some measure of control have changed markedly over time. In the early theological conceptions, human nature was ordained by original divine design. Evolutionism transformed the conception to one in which human nature is shaped by environmental pressures acting on random gene mutations and reproductive recombinations. This nonteleological process is devoid of deliberate plans or purposes.

The symbolic ability to comprehend, predict, and alter the course of events confers considerable functional advantages. The evolutionary emergence of language and abstract and deliberative cognitive capacities provided the neuronal structure for supplanting aimless environmental selection with cognitive agency. Human forebears evolved into a sentient agentic species. Their advanced symbolizing capacity enabled humans to transcend the dictates of their immediate environment and made them unique in their power to shape their circumstances and life courses. Through cognitive self-guidance, humans can visualize futures that act on the present, order preferences rooted in personal values, construct, evaluate, and modify alternative courses of action to secure valued outcomes, and override environmental influences.

The present chapter addresses the issue of free will from the agentic perspective of social cognitive theory (Bandura, 1986, 2006). To be an agent is to influence intentionally one's functioning and the course of environmental events. People are contributors to their life circumstances not just products of them. In this view, personal influence is part of the determining conditions governing self-development, adaptation, and change.

There are four core properties of human agency. One such property is *intentionality.* People form intentions that include action plans and strategies for realizing them. Most human pursuits involve other participating agents so there is no absolute agency. They have to negotiate and accommodate their self-interests to achieve unity of effort within diversity. Collective endeavors require commitment to a shared intention and coordination of interdependent plans of action to realize it (Bratman, 1999). Effective group performance is guided by collective intentionally.

The second feature involves the temporal extension of agency through *forethought*. This includes more than future-directed plans. People set themselves goals and anticipate likely outcomes of prospective actions to guide and motivate their efforts anticipatorily. A future state has no material existence, so it cannot be a cause of current behavior acting purposefully for its own realization. But through cognitive representation, visualized futures are brought into the present as current guides and motivators of behavior. In this form of anticipatory self-guidance, behavior is governed by visualized goals and anticipated outcomes rather than being pulled by an unrealized future state. The ability to bring anticipated outcomes to bear on current activities promotes purposeful and foresightful behavior. When projected over a long time course on matters of value, a forethoughtful perspective provides direction, coherence, and meaning to one's life.

The third agentic property is *self-reactiveness.* Agents are not only planners and forethinkers. They are also self-regulators. Having adopted an intention and action plan, one cannot simply sit back and wait for the appropriate performances to appear, as Searle (2003) notes in his analyses of the explanatory gap. Agency thus involves not only the deliberative ability to make choices and action plans, but the ability to construct appropriate courses of action and to motivate and regulate their execution. This multifaceted self-directedness operates through self-regulatory processes in the explanatory gap to link thought to action (Bandura, 1991a; Carlson, 2002).

The fourth agentic property is *self-reflectiveness.* People are not only agents of action. They are self-examiners of their own functioning. Through functional self-awareness, they reflect on their personal efficacy, the soundness of their thoughts and actions, and the meaning of their pursuits, and they make corrective adjustments if necessary (Bandura, 1986). The metacognitive capability to reflect upon oneself and the adequacy of one's thoughts and actions is the most distinctly human core property of agency.

Much of the theorizing about human self-regulation (Carver & Scheier, 1981; Lord & Levy, 1994) is founded on Powers' (1973) control theory, which is an outgrowth of the cybernetic model of how mechanical devices are self-regulating via negative feedback. The principal driving force is the negative feedback loop. In this regulatory process, deviations from a programmed state detected by a sensor automatically triggers activity that drives the system toward the programmed state to maintain equilibrium in the face of environmental perturbations. The cybernetic system embodies a hierarchy of interconnected feedback loops with upper level loops providing the reference signals that serve as goal settings for subordinate loops. The applicability of robotic self-regulating models to human self-management is critically addressed elsewhere in some detail (Bandura & Locke, 2003; Locke, 1994).

Humans operate as an open, proactive system rather than solely as a reactive cybernetic one (Bandura, 1991a, 1997). They motivate themselves by *discrepancy production* not just discrepancy reduction. They adopt goals and standards that create a state of disequilibrium and then enlist the strategies and effort required to realize them. After attaining the standard they have been pursuing, those with a high sense of efficacy generally set themselves further challenges that create new disequilibrating discrepancies to be mastered. However, goal adjustments do not follow a neat pattern of ever-rising standards following accomplishment, nor are individuals driven automatically to reduce disparity between sensed feedback and inner referent.

Consider the complexity of self-regulative agency. People act proactively in choosing and changing the goals they aim for: They juggle multiple goals and often have to choose between conflicting ones, respond in a variety of possible ways to performance shortfalls, set their slate of options for serious consideration based on judgments of their efficacy, process feedback through their knowledge base and preconceptions, devise functional strategies, override prepotent influences that divert one from a chosen pursuit, and engage in a lot of self-reflective metacognitive activity concerning the adequacy of their self-efficacy appraisals, operative strategies, adopted goal challenges, and outcome expectations. Moreover, they must manage stressors, self-debilitating ideation, and affective self-evaluative reactions to their performances that can undermine self-regulatory efforts.

In short, people have to navigate through complex environments of innumerable variations, novelties, ambiguities, and unpredictability. Much of this environment involves dynamic changes requiring adaptive flexibility in multi-agent transactions in which the participants are both actors and acted upon. They not only can observe what they are doing, but do something to affect the course of events. In even more consequential exercise of agentic capability, individuals create environments not simply react to them in preprogrammed ways. Being a self-governing human is a quite different matter from being a self-regulating thermostat.

ORIGINS OF PERSONAL AGENCY

The newborn arrives without any sense of selfhood and personal agency. Agentic capabilities must be socially constructed through transactional experiences with the environment. The developmental progression of personal agency proceeds from perceiving causal relations between environmental events, through understanding causation via action, and finally to recognizing oneself as the agent of the actions. Infants exhibit sensitivity to causal relations between environmental events even in the first months of life (Lent, 1982; Mandler, 1992). They most likely begin to learn about action causation through repeated observation of contingent occurrences in which the actions of others make things happen. They see inanimate objects remain motionless unless manipulated by others (Mandler, 1992). Moreover, infants personally experience the effects of actions directed toward them, which adds salience to the causative functions of actions.

Recognition of action causation is socially enhanced by linking outcomes closely to infants' actions, by using aids to channel infants' attention when there is a temporal disconnect between their actions and the outcomes they are producing, and by heightening the salience and functional value of the outcomes (Millar, 1972; Millar & Schaffer, 1972; Watson, 1979). As infants begin to develop behavioral capabilities, they not only observe, but directly experience that their actions make things happen. With the development of representational capabilities, infants can begin to learn from probabilistic and more distal outcomes they bring about by their actions.

Development of a sense of personal agency requires more than simply producing effects by actions. Infants acquire a sense of personal agency when they recognize that they can make things happen and they regard themselves as agents of those actions. This additional understanding of oneself as the doer extends the perception of agency from action causality to personal causality. The differentiation of one's own actions as distinct from those of others is the product of a more general process of the construction of an agentic self. Proprioceptive feedback from one's activities and self-referent information from visual and other modalities in transactions with the environment aid in the early perception of an experiential person. Personal effects resulting from self-directed actions further identify oneself as the recipient experiencing the effects. Thus, if touching a hot object brings pain, feeding oneself brings comfort, and entertaining oneself with manipulable objects generates enjoyment, such self-produced outcomes foster recognition of oneself as an agent. One becomes differentiated from others through rudimentary dissimilar experiences. If stubbing one's toe brings pain, but seeing others stub their toe brings no personal pain, one's own activity becomes distinguished from that of other persons.

The construction of personhood is not entirely a matter of private reflection on one's experiences. There is a social aspect to this process. As infants mature and acquire language, those around them refer to them by personal names and treat them as distinct persons. With the development of language, social self-referent labeling accelerates self-recognition and development of self-awareness of personal agency. By about 18 months, infants have self-referent verbal labels and apply them only to pictures of themselves (Lewis & Brooks-Gunn, 1979). They differentiate themselves from others in their verbal labeling. As they become increasingly aware that they can produce effects by their actions, by about 20 months, they spontaneously describe themselves as agents of their actions and their intentions as they engage in activities (Kagan, 1981). Before long, they begin to describe the psychological states accompanying their actions. Based on their growing personal and social experiences, they eventually form a symbolic representation of themselves as a distinct person capable of making things happen.

There is also a great deal of intentional guidance in fostering infants' agentic capabilities (Heckhausen, 1987; Karniol, 1989; Papousek & Papousek, 1979). Parents create highly noticeable proximal effects of infants' actions, segment activities into manageable subskills, and provide infants with objects within their manipulative capabilities that enable them to produce effects by their actions. They set challenges for their infants just beyond the infants' existing competencies. They adjust their level of assistance across phases of mastery, offering explicit guidance in earlier phases of skill acquisition but gradually withdrawing aid as infants become more competent in mastering tasks on their own. These types of enabling strategies are highly conducive to the development of personal agency during the initial years of life.

The self is the person, not a homunculan overseer that resides in a particular place and does the thinking and acting. Personhood embodies one's physical and psychosocial makeup with a personal identity and agentic capabilities that operate in concert through a variety of special-purpose biological systems. Although the brain plays a central role in psychological life, personhood does not reside solely in the brain, any more than the heart is the sole place of circulation (Schechtman, 1997). For example, the musculature of a gymnast honed through countless hours of practice is part of the self but not solely of the brain. Transplanting the brain of an extraordinary gymnast into an octogenarian's body will not produce a self as a dazzling gymnast as a single organ view would imply.

Nor are there multiple independent selves. Human behavior is socially situated, highly contextualized, and conditionally manifested. Adaptive functioning requires both appropriate generalization in the face of bewildering situational variation and perceptive discrimination to avoid dysfunctional overgeneralization. People, therefore, vary in their behavior conditional on circumstances that reflect the diverse aspects of their lives. They wrestle with conflicting goals and courses of action. But these are instances of the same being doing different things under different life conditions, not different selves doing their separate things. Positing multiple selves plunges one into deep philosophical waters. It requires a regress of selves to a presiding superordinate self who selects, and manages the collection of selves, for selected purposes. Given but a single body, the choices finally made and the execution of a chosen course of action requires singleness of agency. The fragmentation of agency into multiple selves poses additional conceptual problems. Once you start fractionating the self, where do you stop?

Social cognitive theory also calls into question conceptions positing a duality of self as agent and as object in self-reflectivity. This seeming ontological separation involves shifting the perspective of the same agent rather than partitioning an self. The shift in perspective does not transform one from an agent to an object. One is just as much an agent reflecting on oneself as in acting on the environment. There is no reified self behind the reflecting.

Ismael (2007) specifies processes governing the synchronic unity of the self-representational system in action. The unifying activity includes construal and integration of information from diverse sources into a single voice through the mind-set of personal experiences. To add to the complexity of the integrating process, some of the information is potentially conflicting. This analysis provides added value because it extends beyond the individual level to the achievement of unity among different constituents within a social system. The latter level of analysis is especially relevant to the exercise of collective agency in which unity is formed from social diversity in the pursuit of common purpose.

Identity formation is an important aspect of human agency. Personal identity refers to a sense of individuality and one's self-characterization. It affects how people structure their lives and relate to the everyday world around them. The psychological issues of interest in self-representation center on the organization and continuity of personal identity in the midst of notable changes over time and across different spheres of life. The transactions of everyday life also require a distinctive social identity that matters in how one is treated.

The continuity of personal identity resides more in psychological factors and the experiential continuity of one's life course than in physical constancy. An amnesic remains the same physically, but has no sense of personal identity. Identity is preserved in memories that give temporal coherence to life, in the connectedness of human relationships and one's life work over time, and in continuance of belief and value commitments that link the present to the past and shape the future.

Our theories place heavy emphasis on phenomenological continuity. In social cognitive theory, personal identity is also rooted in agentic continuity. People not only construe themselves as a continuing person over different periods in their lives; through their goals, aspirations, social commitments, and action plans, people project themselves into the future and shape the courses their lives take (Korsgaard, 1996). In short, they agentically construct continuities.

Continuity in personal identity is not solely a product of an intrapsychic autobiographical process that preserves a sense of personhood over time. Others perceive, socially label, and treat one as the same person over the course of life despite physical changes. Personal identity is partially constructed from one's social identity as reflected in how one is treated by significant others. In keeping with the model of triadic reciprocal determination, an enduring personhood is the product of a complex interplay of personal construal processes, agentically constructed continuity, and influences from the social reality in which one lives.

MODES OF AGENCY

Social cognitive theory distinguishes among three modes of agency: individual, proxy, and collective efficacy. In personal agency exercised individually, people bring their influence to bear on their own functioning and on environmental events. However, in many spheres of functioning, people do not have direct control over conditions that affect their lives. They exercise proxy agency through socially mediated influence. They do so by influencing others who have the resources, knowledge, and means to act on their behalf to secure the outcomes they desire (Baltes, 1996; Brandstädter & Baltes-Gotz, 1990; Ozer, 1995). Children work through parents to get what they want, marital partners through spouses, employees through labor unions, and the general public through their elected officials. However, people often turn to intermediaries in areas of functioning in which they can exercise direct control but choose not to because they have not developed the competencies to do so, they believe others can do it better, or they do not want to saddle themselves with the task demands, stressors, and onerous responsibilities that personal control requires. This socially mediated mode of agency introduces other players and time lags between one's goals and intentions and attainment of desired behavioral outcomes.

People do not live their lives in individual autonomy. Many of the things they seek are achievable only by working together through interdependent effort. In the exercise of collective agency, they pool their knowledge, skills, and resources, and act in concert to shape their future (Bandura, 2000). In this multiagent mode of collective agency, participants have to achieve unity of effort for common cause within diverse self-interests and coordination of distributed subfunctions across a variety of individuals. The distinctive blend of individual, proxy, and collective agency varies cross-culturally. But everyday functioning relies on all three forms of agency to make it through the day, wherever one lives.

A resilient sense of personal and collective efficacy has generated functional value in the exercise of each of these modes of agency in diverse cultural systems (Bandura, 2002). Being immobilized by self-doubt in one's capabilities to produce effects and believing in the futility of effort have little evolutionary advantage. However, the ways in which efficacy beliefs are developed and structured, how they are exercised, and the purposes to which they are put cross-culturally. In short, there is cultural commonality in basic agentic capacities and mechanisms of operation, but diversity in the culturing of these inherent capacities. In this dual-level analysis, universality is not incompatible with manifest cultural plurality, Kluckhohn and Murray summarized eloquently the blend of universality, commonality, and uniqueness of human qualities. Every person is in certain aspects like all other people, like some other people, like no other person (as cited in Muñoz & Mendelson, 2005).

TRIADIC RECIPROCAL DETERMINATION

People do not operate as autonomous agents. Nor is their behavior wholly determined by situational influences. Rather, human functioning is a product of a reciprocal interplay of intrapersonal, behavioral, and environmental determinants (see figure). This triadic interaction includes the exercise of personal influence as part of the determining conditions (Bandura, 1986). The notion of "free will"

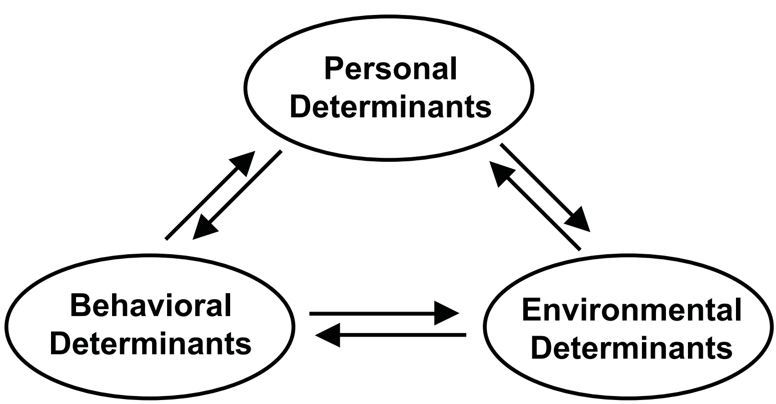

Figure 6.1.

is recast in terms of personal contribution to the constellation of determinants operating within the dynamic triadic interplay.

In the analytic decomposition of triadic determination, different subspecialties of psychology have centered their inquiry on particular segments of the reciprocal interplay.

In the reciprocative relation between intrapersonal and behavioral determinants, people's biological endowments, conceptions, values, goals, and affective states influence how they behave. The natural and extrinsic effects of their actions, in turn, affect their thought processes and affective states. In the reciprocative relation between behavioral and environmental determinants, behavior alters environmental conditions and is, in turn, altered by the very conditions it creates. In the reciprocative relation between intrapersonal and environmental determinants, social influences in the form of social modeling, instructional practices, and various modes of social persuasion alter personal attributes. In the reciprocal impact of this segment, people can affect their environment without saying or doing anything. They elicit reactions from the social environment simply by their physical characteristics, such as their ethnicity, gender, race, age, physical attractiveness, and their socially conferred roles and statuses. The social reactions thus elicited, in turn, affect the recipients' conceptions of themselves and others in ways that either strengthen or reduce the environmental bias.

Psychosocial accounts of human functioning often portray individuals as reactors to environmental events impinging upon them. In the neurophysiological quest for the localization and neural circuitry underlying cognitive activities and affective reactions, individuals are transported in a prone position into a neuroimaging device in which they are greeted with stimuli to which they have to react instantly. Such an arrangement allows little leeway for deliberative proactive control of action, much of which must be psychosocially negotiated and temporally regulated in everyday life. Social cognitive theory distinguishes among three different types of environments: imposed, selected, and created. The imposed physical and sociostructural environment impinges on people whether they like it or not. They have little control over its presence, but they have some latitude in how they construe it and react to it. However, for the most part, the environment is only a potentiality that does not come into being until it is selected and actualized by the actions that are taken. This constitutes the selected environment. Under the same potential environment, some people take advantage of the opportunities it provides and its rewarding aspects. Others get themselves enmeshed mainly in its debilitating and aversive aspects. We are all acquainted with problem-prone individuals who, through their irksome conduct, breed negative social climates wherever they go. Others are equally skilled at bringing out the best in those with whom they interact. People also construct physical and social environments that enable them to exercise some measure of control over their lives.

Gradations of environmental controllability require increasing levels of personal agency. Use of imposed environments calling for simple actions with scarcely any options in an invariant environment limits the generalizability of findings to common life conditions in which people have some leeway to select and construct environments and shape the course of events. We will revisit this issue when considering neurophysiological studies of the cognitive control of action.

Many factors enter into the production of given outcomes. Because of the multiplicity of interacting influences, the same factor can be part of different blends of codetermining conditions (Bandura, 1986). Moreover, reciprocality does not mean symmetry in the strength of bidirectional influences. Nor is the patterning and strength of mutual influences fixed in reciprocal determination. The relative magnitude of the personal contribution to the codetermination within the triadic system varies depending on the level of agentic personal resources, types of activities, and situational circumstances. Nor does reciprocality mean holistic simultaneous influence (Bandura, 1983). A given determinant and its reciprocal effects do not spring forth concurrently. It takes time for a determinant to exert its influence. Because the triadic determinants do not operate simultaneously as an unravelable holism, the temporal lags between events enable one to clarity how they function interactively.

Most human commerce with the environment is socially situated in interpersonal transactions. This is becoming increasingly so with the revolutionary advances in communication technologies. People are now spending much of their time in the cyberworld, where they not only have a vast array of options to choose from at their fingertips, but they are posting their own constructions in this virtual environment for response from others. In an interpersonal transaction, individuals are each other's environments. Consequently, the status of psychological constructs change in the flow of social embeddedness. Person A becomes the agent acting on the environment (Person B) if one enters the transactional analysis on the A side. But Person B's status changes from an environment to an agent acting on the environment (Person A) if one enters on the B side one step later in the ongoing transaction. Thus, the same event can change from an agentic influence to a behavioral expression, and to an environmental outcome, depending arbitrarily on different entry points in the ongoing transaction between the individuals involved.

In addition to taking a hand in shaping their external environment, people live in a psychic environment largely of their own making. The self-management of inner life is also part of the agentic process. This line of inquiry is providing new knowledge into people's capability to regulate their own thought patterns and affective states by enlisting a variety of cognitive and behavioral strategies (Bandura, 1997; Lazarus & Folkman, 1984; Rosenthal & Rosenthal, 1985; Wegner, 1989). In these efforts, self-regulation of one's consciousness is the object of study. In clinical applications of this knowledge, people improve the quality of their functioning and emotional well-being by exercising control over ruminative perturbing and dejecting thoughts. This control process is well captured in the Chinese proverb, "You cannot prevent the birds of worry and care from flying over your head, but you can prevent them from building a nest in your head." Anderson and his colleagues (Anderson et al., 2004) have identified neural systems underlying the intentional control of consciousness.

Any retrospective causal analysis must include the triadic interplay of determinants rather than a truncated regression solely to the external environmental facet acting autonomously and unidirectionally. Humans are not like billiard balls propelled solely by forces external to them. Billiard balls cannot change the shape of the table, the size of the pockets, or intervene in the paths they take, or even decide whether to play the game at all. In contrast, humans not only think, but, individually and collectively, shape the form those external forces take and even determine whether or not they come into play. Murray Gell-Mann, the physicist Nobelist, underscored the influential role of the personal determinants when he remarked, "Imagine how hard physics would be if particles could think" (Gruman, 2006). Environmental influences, of course, contribute to the development of personal attributes. But here, too, they are the product of the triadic interplay of personal, behavioral, and environmental factors rather than created by an autonomous environmental force. In short, personal influence is a significant player, not only in the proximate determining conditions, but in the prior chain of determination as well.

INTERPLAY OF HUMAN AGENCY AND SOCIAL STRUCTURE

Human functioning is rooted in social systems. Therefore, personal agency operates within a broad network of sociostructural influences. These social systems are devised to organize, guide, and regulate human affairs in diverse spheres of life by authorized rules, sanctions, and enabling resources (Giddens, 1984). Social systems do not arise by immaculate conception. Social cognitive theory rejects a duality of human agency and a social structure as a reified entity disembodied from individuals. Social systems are the product of human activity. The authorized rules and practices of social systems, in turn, influence human development and functioning. However, in the dynamic interplay within the societal rule structures, there is a lot of personal variation in the interpretation, adoption, enforcement, circumvention, and opposition to societal prescriptions and sanctions (Burns & Dietz, 1992).

It has been shown that sociostructural influences operate, to a large extent, through psychological mechanisms to produce behavioral effects (Bandura, 1995, 1997; Baldwin, Baldwin, Sameroff, & Seifer, 1989; Elder, 1995). However, as previously noted, in agentic transactions people are producers of their lived environment not merely a personal conduit through which sociostructural influences operate.

TRIADIC DETERMINATION AND FREEDOM

The exercise of human agency raises the issue of freedom and determinism. Humans are not just reactive to external input in a preprogrammed, robotic way. As noted in the functional properties of human agency and triadic codetermination, intrapersonal influences are significant contributors to the course of events. In a similar vein, Ismael (2006, 2007) builds a strong case that deliberation brings into play a variety of intrapersonal influences that can break the chain of determination from external conditions to action. Moreover, within the triadic codetermination, deliberative thought not only alters the relation between environmental influences and behavioral outcomes, but fosters courses of action that proactively shape the physical and social environments.

When viewed from a social cognitive perspective, freedom is not conceived just passively as the absence of constraints and coercion in choice of action, but proactively as the exercise of self-influence in the service of selected goals and desired outcomes. For example, people have the freedom to vote, but whether they get themselves to vote and the level and form of their political engagement depends, in large part, on the self-influence they bring to bear. In addition to regulating their actions, as previously noted, people also live in a psychic environment. The self-management of one's inner life frees one from unwanted trains of thought. Because personal influence is an interacting part of the determining conditions, freedom is not incompatible with one's actions being determined. People are partial authors of the past conditions that developed them as well as the future course their lives take. Analyses of freedom are typically framed in a contentious dualism pitting determinism against randomness of causation. The model of triadic reciprocity provides a nuanced view in which individuals are contributors to the determining conditions.

The cultivation of agentic capabilities adds concrete substance to abstract metaphysical discourses about freedom and determinism. People who develop their competencies, self-regulatory skills, and enabling beliefs in their efficacy can generate a wider array of options that expand their freedom of action. They are also more successful in realizing desired futures than those with less developed agentic resources (Bandura, 1986). The development of strategies for exercising control over perturbing and self-debilitating ideation is intrapsychically liberating.

There is no absolute freedom. Paradoxically, to gain freedom individuals have to negotiate consensual rules of behavior for certain activities that require some relinquishment of autonomy. Without traffic laws, for example, driving would be chaotic, perilous, unpredictable, and uncontrollable for everyone.

The exercise of freedom involves rights as well as options and the means to pursue them. At the societal level, people institute, by collective action, regulatory sanctions against unauthorized forms of societal control (Bandura, 1986).

The less social jurisdiction there is over certain activities, the greater is the contribution of personal influence to choice of action in those domains. After protective laws are built into social systems, there are certain things that a society may not do to individuals who choose to challenge conventional values or vested interests, however much it might like to. Legal prohibitions against unauthorized societal control create personal freedoms that are realities, not illusory abstractions.

Societies differ in their institutions of freedom and in the number and type of activities that are officially exempted from social control. For example, social systems that protect journalists from criminal sanctions for criticizing government officials are freer than those that allow authoritative power to be used to silence critics or their vehicles of expression. Societies that possess a judiciary independent of other government institutions ensure greater social freedom than those that do not.

AGENTIC MANAGEMENT OF FORTUITY

There is much that people do designedly to exercise some measure of control over their self-development and life circumstances. But there is a lot of fortuity in the courses lives take. Indeed, some of the most important determinants of life paths occur through the most trivial of circumstances. People are often inaugurated into new life trajectories, marital partnerships, and occupational careers through fortuitous circumstances (Austin, 1978; Bandura, 1986; Stagner, 1981). In their insightful volume *The Travels and Adventures of Serendipity*, Merton and Barber (2004) document the workings of fortuitous events in life trajectories.

A fortuitous event in social encounters is an unintended meeting of persons unfamiliar with each other. The physical sciences acknowledge indeterminacy at the quantum mechanical level in the physical world. Fortuitous events introduce an element of indeterminacy in the behavioral sciences. The separate paths have their own determinants, but they are causally unconnected until their intersection, at which point the encounter creates a unique confluence of influences that can alter life courses. The intersection, where the transactions take place, occurs fortuitously rather than by design within the deterministic context (Nagel, 1961). Consider an example of a fortuitous event at an address on the psychology of chance encounters that altered the course of lives (Bandura, 1982). An academic publisher entered the lecture hall as it was rapidly filling up and seized an empty chair near the entrance. He ended up marrying the woman who happened to be seated next to him. With only a momentary change in time of entry, seating constellations would have altered and this intersect would not have occurred. A marital partnership was thus fortuitously formed at a talk devoted to fortuitous determinants of life paths!

A seemingly insignificant fortuitous event can set in motion constellations of influences that change life courses These branching processes alter the continuity and linear progression of life-course trajectories. The profusion of separate chains of events in everyday life provides myriad opportunities for such fortuitous intersects. Even if one knew all of the determinate conditions for particular individuals, one cannot know in advance the intersection of unconnected events. Fortuitous intersects introduce probabilistic uncertainties that complicate long-range predictions of human behavior.

Most fortuitous events leave people untouched, others have some lasting effects, and still others branch people into new trajectories of life. A science of psychology does not have much to say about the occurrence of fortuitous intersects, except that personal proclivities, the types of settings in which one moves, and the types of people who populate those settings make some types of intersects more probable than others. Fortuitous occurrences may be unforeseeable, but having occurred, the conditions they create operate as contributing factors in causal processes in the same way as do prearranged ones. Hence, psychology can advance knowledge on the effects of fortuitous events on life paths. Several lines of evidence identify personal attributes and the properties of the environments into which individuals are fortuitously inaugurated as predictors of the nature, scope, and strength of the impact that such encounters are likely to have on personal lives (Bandura, 1982, 1986).

Fortuity does not mean uncontrollability of its effects. People can bring some influence to bear on the fortuitous character of life. They can make chance happen by pursuing an active life that increases the number and type of fortuitous encounters they will experience (Austin, 1978). Chance favors the inquisitive and venturesome, who go places, do things, and explore new activities. People also make chance work for them by cultivating their interests, enabling beliefs, and competencies (Bandura, 1998). These personal resources enable them to make the most of opportunities that arise unexpectedly. Pasteur put it well when he noted, "Chance favors only the prepared mind." Even that distinguished lay philosopher, Groucho Marx, insightfully observed that people can influence how they play the hand that fortuity deals them, "You have to be in the right place at the right time, but when it comes, you better have something on the ball." Self-development gives people a hand in shaping the courses their lives take. These various proactive activities illustrate the agentic management even of fortuity.

GENETIZATION OF HUMAN BEHAVIOR

We are currently witnessing an extensive genetization of human behavior. Social roles and human practices are increasingly being proclaimed as driven by prehistoric biological programming.

Biology provides the information-processing systems and physical potentialities and sets constraints. But in most spheres of functioning, biology permits a broad range of cultural possibilities. Boyd points out (Dreifus, 2005) that humans evolved in the tropics but hunt seals in the Arctic. Genes did not teach them how to build a kayak; their culture did. As Gould (1987) has correctly observed, the major explanatory dispute is not between nature and nurture as the issue is commonly framed. Rather, the issue in contention is whether nature operates as a determinist that has culture on a "tight leash," as Wilson (1988) contends, or as a potentialist, that has culture on a "loose leash," as Gould (1987) maintains.

Humans have created societies of diverse natures: aggressive and pacific ones, egalitarian and despotic ones, altruistic and selfish ones, individualistic and collectivistic ones, and enlightened and backward ones. Evidence supports the potentialist view. For example, people possess the biological capability for aggressive acts but cultures differ markedly in aggressiveness (Alland, 1972; Gardner & Heider, 1969; Levy, 1969). There are also wide differences in aggression within the same culture (Bandura, 1973). Even entire nations, such as Sweden and Switzerland, have transformed from warring societies to pacific ones. The Swiss used to be the main suppliers of mercenary fighters in Europe. As they transformed into a pacific society, their militaristic vestige was evident only in the plumage of the Vatican guards. For ages, the Vikings plundered other nations. Their ruthlessness was captured in the prayer, "Deliver me, O Lord, from the fury of the Norsemen." After a prolonged war with Russia, the populace rose up and forced a constitutional change. It prohibited kings from starting wars (Moerk, 1995). This political act promptly transformed a warring society into a peaceful one. Sweden is now a mediator for peace among warring nations. Cultural diversity and the rapid transformative societal changes underscore that the answer to human aggression lies more in ideology than in biology.

GROWING PRIMACY OF HUMAN AGENCY IN THE COEVOLUTION PROCESS

Dobzhansky (1972) reminds us that humans are a generalist species that was selected for learnability and plasticity of behavior, not for behavioral fixedness. Although not limitless, changeability and agentic capability are the hallmark of human nature. Because of limited innate programming, humans require a prolonged period of development to master essential competencies. Moreover, different periods of life present new competency demands requiring self-renewal over the life course to meet the challenges of changing norms and life circumstances. To add to the necessity of changeability, the eras in which people live usher in technological innovations, shifts in socioeconomic conditions, cultural upheavals, devastating wars, and political changes that make life markedly different calling for new adaptations (Elder, 1994). These diverse adaptational changes are cultivated by agentic psychosocial means.

People are not just reactive products of selection pressures served up by a one-sided evolutionism. They are prime players in the coevolution process.

Other species are heavily innately programmed as specialists for stereotypic survival in a particular habitat. In contrast, through agentic action, people devise ways of adapting flexibly to remarkably diverse geographic, climatic, and social environments. They devise ways to transcend their biological limitations. For example, humans have not evolved morphologically to fly but they are soaring through the air and even in the rarified atmosphere of outer space at breakneck speeds despite the inborn constraint. Agentic inventiveness transcended genes and biological design in getting them airborne.

People use their ingenuity to circumvent and insulate themselves from selection pressures. They create devices that compensate immensely for their sensory and physical limitations. They construct complex environments to fit their desires, many of which are fads and fashions that are socially constructed by vigorous marketing practices. They create intricate styles of behavior necessary to thrive in complex social systems. Through social modeling and other forms of social guidance, they pass on to subsequent generations accumulated knowledge and effective practices. They transcend time, place, and distance as they interact globally with the symbolic environment of the cyberworld.

Through contraceptive ingenuity, which disconnected sex from procreation, humans have outwitted and taken control over their evolved reproductive system. They seek sex without procreation rather than strive to propagate their kind in large numbers. They are developing reproductive technologies to separate sex even from fertilization. Through genetic engineering, humans are creating biological natures, for better or for worse, rather than waiting for the slow process of natural evolution. They are now changing the genetic makeup of plants and animals. Unique native plants that have evolved over eons are disappearing as commercial horticulturalists are supplanting them with genetically uniform hybrids and clones. Humans are not only cutting and splicing nature's genetic material, but, through synthetic biology, they are creating new types of genomes. Humans are even toying with the prospect of fashioning some aspects of their own biological nature by genetic design.

The inventive power of human agency is largely ignored in evolutionary accounts of human behavior, especially in the more biologically deterministic views propounded in psychological evolutionism. Given the growing human modifications of evolved heritages and creative circumventing of endowed limitations, the notion in vogue that biological evolution provides the potential and culture can do only so much with it flies in the face of the extraordinary control wielded by inventive human agency. The psychosocial side of coevolution is gaining ascendancy through the agentic power to transform environments and what humans become. In short, we are an agentic species that can alter evolutionary heritages and shape the future.

Social cognitive theory does not dismiss the contribution of genetic endowment to human adaptation and change. On the contrary, this endowment provides the very neuronal structures and mechanisms for the agentic properties that are distinctly human. These include generative thought, symbolic communication, forethought, self-regulation, and reflective self-consciousness. The uniqueness of humans resides in these self-directing and self-transforming capacities. Neither the agentic human ascendance in the coevolution process nor the rapid transformational societal changes it spawns would be possible without the biological endowment of abstract cognitive capabilities. What is disputable is the common practice of attributing human affairs to alleged vestiges of prehistoric conditions that are unknowable. Social cognitive theory highlights the forward-looking impact of our biological endowment, rather than backward-looking conjectures about adaptation to prehistoric conditions. The study of how humans are changing endowed heritages, circumventing biological constraints, and shaping their future through social and technological evolution has greater promise of providing new insights into the diverse patterns of human adaptation in contemporary times than spinning fanciful stories about prehistoric mating patterns in drafty caves.

NONAGENTIC THEORETICAL APPROACHES

In its brief history, psychology has undergone wrenching paradigm shifts. Behaviorists proposed an input → output model linked by an internal conduit that makes behavior possible but exerts no influence of its own on behavior. In this view, human behavior was shaped and controlled automatically and nonconsciously by environmental stimuli. This line of theorizing was eventually put out of fashion by the advent of the computer, which likened the mind to a biological calculator. Creative thinkers filled the internal conduit with symbolic representations, rules, and computational operations. The mind as a symbol manipulator, in the likeness of a linear computer, became the conceptual model for the times. The input → output model was thus supplanted by an input → linear throughput → output model. For decades, the reigning metaphor of human functioning was a linear computational system in which information is fed through a central processor that cranks out solutions nonconsciously according to preordained rules. The architecture of the linear computer at the time dictated the conceptual model of human functioning.

Computerized serial cognitivism was, in turn, supplanted by connectionist models that operate through interconnected, multilayered neuronal-like subsystems working simultaneously in parallel. Sensory organs deliver up information to a multitude of subsystems acting as the mental machinery that processes the inputs and generates a coherent output automatically and nonconsciously out of the fragmentary neuronal activity. The cognitive machinery operating through associated networks does the construing, planning, motivating, and regulating nonconsciously. The inputs from these special purpose miniprocessors have to be integrated and coordinated to be able to act in a purposeful, coherent way. Given the extensive neuronal interconnectedness, this rarely occurs in a single anatomical location. Without a coordinative function, it remains in foggy ambiguity how a decentralized system with miniprocessors doing their own thing can operate as a unified whole in pursuit of selected goals. Moreover, as indicated earlier, people are shapers of their environment, not just information processors of environmental inputs.

Green and Vervaeke (1996) report that originally many connectionists and computationalists regarded their conceptual models as approximations of cognitive activities. They include representations of goals and other internal states in the regulation of human behavior (Miller & Cohen, 2001). But some connectionists have become eliminative physicalists, likening cognitive factors to the phlogiston of yesteryear. In this view, people do not act on beliefs, goals, aspirations, and expectations. Rather, activation of their network structure at a subpersonal level makes them do things. In a critique of eliminativism, Greenwood (1992) notes that cognitions are contentful psychological factors whose meaning does not depend on the explanatory propositions in which they figure. As for the phlogiston analogy, this mysterious substance neither had any evidential basis nor explanatory or predictive value. In contrast, cognitive factors do quite well in predicting human behavior and guiding effective interventions. To make their way successfully through a complex world full of hazards, people have to make good judgments about their capabilities, anticipate the probable effects of different events and courses of action, size up sociostructural opportunities and constraints, and regulate their behavior accordingly. These belief systems are a working model of the world that enables people to achieve desired outcomes and avoid untoward ones.

Forethoughtful, regulative, and reflective capabilities are vital for survival and human progress. The theorizing at the psychostructural level of complexity and the accompanying verified knowledge of psychosocial regulation of behavior cannot be cavalierly dismissed as merely folk psychology of the phlogiston variety. Agentic factors that are explanatory, predictive, and of demonstrated functional value may be translatable and modeled in another theoretical language but are not eliminatable (Rottschaefer, 1985, 1991). Progress in the understanding of human behavior is better served by clarifying links across levels of complexity than by reductive dismissal of verified principles operating at the higher level.

The various nonagentic theories differ in what they place in the mediating system, whether it includes determinative functions, and the forms they take. The theories posit a noncausal conduit in radical behaviorism, a linear central processor in computerized cognitivism, and interconnected, neuronal-like subunits in parallel distributed connectionism. But they share the same bottom-up driven causation: Input → Throughput → Output. In each of these models, the environment acts on the biological machinery that generates the output automatically and nonconsciously. In the more radical forms of theorizing, what goes on inside a human agent is not subject to deliberative conscious control.

In agentic theories, cognitive factors in the form of self-views, beliefs, goals, expectations, and mind-sets influence how bottom-up inputs are encoded, organized, and remembered. These are internally generated inputs in the top-down regulation of behavior. Windmann (2005) reviews findings from diverse lines of research showing that higher cognitive processes, operating principally in prefrontal cortical sites, affect how bottom-up input information is processed in perceptual and memory performance, affective reactivity, and decision making. Research using single-neuron recording is shedding new empirical light on the dynamic interplay between bottom-up input information and top-down regulation by neural representations of the outside world (Naya, Yoshida, & Miyashita, 2001; Tomita, Ohbayashi, Nakahara, & Miyashita, 1999).

Nonagentic conceptions strip humans of agentic capabilities, a functional consciousness, and a personal identity. As Harré (1983) noted in his analysis of computationalism, it is not sentient individuals, but their subpersonal parts, that are orchestrating activities nonconsciously. In actuality, however, people act on the environment. They create it, preserve it, transform it, and even destroy it, rather than merely react to it as a given. As will be shown later, these outcomes involve a socially embedded interplay between the exercise of personal agency and environmental influences.

It should be noted in passing that to elude a self-negating predicament, proponents of nonagentic theories implicitly exempt themselves from their theories of how other folks behave. For example, Skinner argued that humans are shaped and controlled by environmental forces under the illusion that they influence events. But he exhorted people to become operant agents shaping their society by applying his operant-conditioning methods. Radical postmodernists, who emphasize fragmentation and relativity, argue authoritatively for the correctness of their view that there is no one correct view. Eliminative physicalists contend that people's behavior is orchestrated unconsciously by their neural networks while mistakenly believing that they are exercising control. But eliminativists do not portray their own cherished treatises as the product of automatic writing by their neural network under illusory personal authorship.

PHYSICALISTIC THEORY OF HUMAN AGENCY

The mind is the embodiment of conscious cognitive states and processes rather than exists as something apart from the brain. The Cartesian substance dualism, which is almost universally rejected by cognitive scientists, forces one to address the formidable explanatory challenge for a physicalistic theory of human agency and a nondualistic cognitivism. Cognitions are high-level cerebral events involving deliberative, reflective, referential, and evaluative processes, not immaterial entities. It is not a hyphenated mind-body structure involving anatomically separate physical entities acting on each other in a Cartesian physicalism.

Rather, it involves highly interconnected brain systems serving different functions subject to higher level control operating within the same material entity. In short, mind is part of hierarchically embedded systems not a separate entity acting on the body. The advanced symbolizing capacity, neuronally distributed and richly interconnected to diverse sensory and motor systems, provides humans with the means to function as mindful agents.

Cognitive regulation operates at the higher level brain structures. Miller and Cohen (2001) review a growing body of neurophysiological and neuropsychological research verifying top-down cognitive regulation of rule-based and goal-directed behavior. The prefrontal cortex plays an especially influential role in cognitive regulation through its dense connections with a wide range of sensory and motor systems, and limbic structures governing affective and motivational processes. Research on neuromotor prostheses with individuals who have lost sensory and motor functions in paralyzed limbs provides a novel way of verifying deliberative regulation of action (Hochberg et al., 2006). A sensor, implanted in the motor cortex, delivers brain signals to a computer connected to a robotic arm enables a person to use thoughts to guide a prosthetic hand and robotic arm to perform the cognized actions. This type of research can add greatly to our understanding of the organization and temporal regulation of actions by goals and intentions in a top-down forethoughtful way.

Some of the neurophysiological studies of self-relevant cognitive activities include relocation of working memory and attentional resources between external events and self-generated thoughts in the service of self-regulation, and selective disengagement from prepotent external stimuli. These cognitive activities are accompanied by changes in activation mainly in prefrontal and parietal regions of the brain (Gusnard, 2005). Such lines of research are beginning to delineate some of the neurological structures essential for a functional personhood and the self-referential neural circuitry through which it is exercised. The brain is trained during socialization and identity formation in self-representation that is consequential in its operation. Thus, the neural circuitry subserving one's own intentions differs from the circuitry accompanying recognition of others' intentions (Becchio, Adenzator, & Bara, 2005). Given that the brain acts in terms of self-representation, self-referent processes warrant serious study rather than dismissal as a homunculan contrivance because some folks view the *self* not as the person, but as a reified manager residing in a particular place. For reasons given earlier, the neurophysiological bases of agentic activities will not be confined to an anatomically unique structure christened as a *self*. As the preceding studies show, neuroscience has moved beyond expunging an autonomous homunculus to research that advances understanding of the nature and function of higher level cognitive control, and the role played by self-referent processes in human functioning.

There is a difference between a reified self lodged in a control center and a self-representational system that comprises functional properties developed through extensive learning and socialization experiences. These include, among other properties, a personal identity, appraisal of personal capabilities, goals linked to values that give purpose and direction to one's activities, discerned conditional relations that permit forethoughtful actions, and self-reactive capabilities rooted in personal standards of merit, responsibility, and morality. Life experiences are processed through this self-referential context rather than processed impartially as though one were devoid of any personal investment. The difference between humans as a self-representational system versus simply a bundle of associative networks operating subpersonally is illustrated in the impact of failure experiences on depressive dysfunctions. In the latter view, failures activate across the associative network memory of past failings with their accompanying negative affect. In the former view, failures activate a self-representation as an inefficacious and unworthy person (Teasdale, 1988). The detrimental representation gives rise to depressive dysfunctions through its impact on cognitive, motivational, affective, and decisional processes (Bandura, 1997).

Cognitive capabilities provide individuals with the means to function as mindful agents. Cognitive activities manifested in consciousness not only provide the means to make life personally manageable but worth living. Consciousness encompasses multiple functions that reflect the difference between being conscious of an activity and consciously engaging in purposeful activity (Korsgaard, 1989). It includes a nonreflective and reflective awareness facet, and a conceptual functional facet operating mainly through the linguistic medium. The functional aspect of consciousness involves purposeful accessing and deliberative processing of information for selecting, constructing, regulating, and evaluating courses of action. This is achieved through intentional recruitment and productive use of semantic and pragmatic representations of activities, goals, and other envisioned future events.

In his discerning analysis of experienced cognition, Carlson (1997, 2002) documents the central role that consciousness plays in the cognitive regulation of action and the flow of ideational events. There have been some attempts to reduce consciousness to an epiphenomenal by-product, to an executive subsystem in the information processing machinery, or to an attentional aspect of information processing. A currently popular solution for the consciousness problem posits an interpretive module that concocts fanciful stories about personal influence over one's behavior that is said to be actually subpersonally determined by low-level neural activity. In this conception, one can have consciousness but need not worry about its functional value because it is dismissed as merely epiphenomenal. Some eliminative physicalists simply redefine out of existence this nettlesome phenomenon that keeps intruding into the cognitive machinery. In the subpersonal accounts of consciousness, there is no experiencing person conceiving of ends and acting purposefully to attain them. These reductive accounts strip the prime features of humanness such as subjectivity, deliberative self-guidance, and reflective self-reactiveness. Without a phenomenal and functional consciousness, people are essentially higher level automatons undergoing actions devoid of any subjectivity or conscious control. Nor do such beings possess a meaningful phenomenal life or a continuing personal identity derived from how they live their life and reflect upon it.

Consciousness is an emergent brain activity with higher-level control functions rather than simply an epiphenomenal by-product of lower level processes. It poses daunting explanatory challenges. Why do humans have a consciousness that operates as a reigning symbolic environment during virtually all of their waking life, if it does nothing? How mind arises from lower level brain processes remains an intractable problem. If the neuronal processes of common activities were automatically reflected in consciousness, it would be hopelessly cluttered with mind-numbing mechanical contents that foreclose any functionality. In driving a car, for example, one's consciousness is filled with thoughts of other matters rather than simply mirroring epiphenomenologically the ongoing neuronal mechanics of driving. What governs which events from among the profusion of ongoing neuronal activities make it into consciousness? We know that people can regulate what inhabits their consciousness through the use of cognitive and behavioral strategies (Bandura, 1997; McCaul & Malott, 1984; Wegner, 1989). There is much work to be done to clarify how lower level brain processes are intentionally recruited in top-down cognitive control to realize given purposes. In this daunting research agenda, researchers have to fend off the specter of homunculan causation.

Social cognitive theory subscribes to a model of emergent interactive agency (Bandura, 1986, 1999a). Cognitive processes are emergent brain activities that exert determinative influence. In emergence, constituent elements are transformed into new physical and functional properties that are not reducible to the elements. For example, the novel emergent properties of water, such as fluidity and viscosity, are not simply the combined properties of its hydrogen and oxygen microcomponents (Bunge, 1977). Through their interactive effects, the constituents are transformed into new phenomena. Van Gulick (2001) makes the important distinction between emergent characteristics of new phenomena and emergent causal powers over events at the lower level.

In the metatheory of cognitive functionalism enunciated by Sperry (1991, 1993), the patterns of neural activities characterizing interpretive and deliberative thought processes have a downward regulatory function over lower level brain processes that lead to action. It will be recalled from the earlier discussion that the evolutionary emergence of a language processing system provided the neuronal structure for the development of a conscious agentic species. Most human thinking operates through language drawing on a vast knowledge base. The core agentic capabilities of intentionality, forethought, self-reaction, and self-reflection operate as hierarchically organized determinants. Their structural and functional properties are central to the exercise of human agency.

SECOND-ORDER CONTROL OF NEUROPHYSIOLOGICAL PROCESSES

In acting as agents, individuals obviously are neither aware of nor directly control their neuronal mechanisms. Rather, they exercise *second-order control*. They do so by intentionally engaging in activities at the macrobehavioral level known to be functionally related to given outcomes. In pursuing these activities, over which they can exercise direct control, they shape their neural circuitry and enlist subpersonal neurophysiological events subserving their chosen pursuits. For purposes of illustration, consider the following analogy. In driving an automobile to a desired place, the driver engages in coordinated acts of shifting gears, steering, manipulating the gas pedal, and applying brakes. The assemblage of auto subsystems provides the intricate operational mechanisms, but they require distinctive higher order activation and regulation. In this multilevel interplay, the acts of driving, which the driver controls directly, regulate the mechanical machinery to get safely to where the driver wants to go. But the driver has neither awareness nor understanding of the correlative microcombustion, transmission, steering, and braking processes subserving the driver's purposes. The deliberate planning of where to go on a trip, what route to take, where to stay, what to do when one gets there and securing reservations for these diverse activities far in advance requires considerable proactive top-down cognitive regulation. The temporal structuring of behavior sets the course for one's activities. Proximal self-regulation provides the guides, strategies, and motivators in the here and now to get to where one is going (Bandura, 1991a). Having constructed a vacation plan, travelers cannot sit back and wait for lower level sensory-motor activity to consummate the vacation arrangements unconsciously.

Consider the second-order control over the intricate neurophysiological machinery. Individuals obviously do not intentionally direct their atrial and ventricular cardiac muscle fibers to fire and their aortic and pulmonary valves to open and close. However, by intentionally engaging in an exercise routine and controlling their activity level, they can enhance their cardiac function and regulate their heart rate without having the foggiest idea of how they indirectly recruited, by their intentional actions, the subserving neurophysiological mechanisms. They can also intentionally speed up and slow down their heart rate by generating frightening and tranquilizing thoughts. In short, enactments of functional activities at the controllable macrobehavioral level serve as the means for agentic recruitment of the subserving brain mechanisms at the microneural level. Framing the issue of conscious cognitive regulation in terms of direct control over the neurophysiological mechanics of action production casts the issue in the wrong terms at the wrong level of control.

Much of the psychological theorizing and research are devoted to verifying functional relations between actions and outcomes and the governing sociocognitive processes. Because individuals have no awareness of their brain processes does not mean that they are just quiescent hosts of automata that dictate their behavior. Neuroimaging can shed light on the neural mechanisms of cognitive control and how controllable agentic action indirectly develops functional neuronal structures and orchestrates the neurodynamics for selected purposes.

PROACTIVE AGENTS VERSUS ONLOOKING HOSTS

One must distinguish between understanding how the biological machinery works in implementing cognitive algorithms by nervous systems, and how the biological machinery is orchestrated agentically for diverse purposes. To use an analogy, knowing the laws of chemistry and physics on how a television set produces images does not explain the endless variety of creative programs it displays. The creative neuronal activation must be distinguished from the neuronal mechanical action production.

People are contributors to their activities not just onlooking hosts of subpersonal networks autonomously creating and regulating their performances. People conceive of ends and work purposefully to achieve them. They are agents of experiences not just undergoers of experiences. In their transactions with their environment, cognitive agents are generative, creative, proactive, and reflective, not just reactive to external input. The sensory, motor, and cerebral systems are tools people use to accomplish the tasks and goals that give meaning, direction, and satisfaction to their lives (Bandura, 1997; Harré & Gillet, 1994). These tools do not come fully prestructured for complex skills. An aspiring violinist, for example, has to practice tenaciously to train the brain, build muscular strength and dexterity, and hone sensory acuity to realize a virtuoso performance. For example, the remarkably versatile brain has to be trained to execute the pyrotechnical wizardry of a Paganini violin concerto. It takes extensive guided practice to configure the brain circuitry to realize this achievement. Purposed action is not the only way to train the brain, however. Thoughts change the brain by cognitive practice in much the same way as does physical practice (Pascal-Leone, et al., 1995). Although the performance gains are lower, prior cognitive practice reduces the time needed to learn a skill by physical practice. There is much excitement about how the brain regulates behavior to the neglect of how individuals train the brain to serve desired purposes.

Research on brain development underscores the influential role that agentic action plays in shaping the functional structure of the brain (Diamond, 1988; Kolb & Whishaw, 1998). It is not mere exposure to stimulation but agentic action in exploring, manipulating, and influencing the environment that counts. By regulating their motivation and activities, people produce the experiences that form the functional neurobiological substrate of symbolic, social, psychomotor, and other skills. An agentic perspective fosters lines of research that can provide new insights into the social and behavioral shaping of brain function. This is a realm of inquiry in which psychology can make unique contributions to the biopsychosocial understanding of human development, adaptation, and change. In nonreductive physicalism, all psychological phenomena have a physical basis. Research from an agentic perspective, however, goes beyond the anatomical localization and brain circuitry subserving human activities to advancing knowledge on brain development and its functional organization by behavioral means (Dawson, Ashman, & Carver, 2000).

NONREDUCTIVE PHYSICALISM

A theory of human agency raises the question of reductionism. One must distinguish among three different forms of reductionism (Ayala, 1974). In ontological reductionism, which is almost invariably adopted, mental events are physical states and processes not disembodied immaterial ones. Epistemological reductionism, across specialized scientific disciplines, contends that the laws governing higher level psychosocial phenomena are ultimately reducible to the laws operating at atomic and molecular levels. Methodological reductionism maintains that research on rudimentary processes will explain psychosocial phenomena at higher levels of complexity. In the heyday of behaviorism, for example, elementary processes were explored with animal analogues using mainly rats and pigeons. The knowledge gained through the study of rudimentary processes is generalizable to some aspects of human functioning, but there are limits as to what it can tell us about the complex human capacity for abstraction and symbolic thinking or the workings of societal systems.

It is the epistemological form of reducibility that is most in contention. The major argument against it is that each level of complexity—physical, chemical, biological, psychological, and social structural—involves emergent new properties that are distinct to that level. In this transformative process, the simpler constituent elements produce qualitatively new phenomena through their interactive effects. The new phenomena at each level of functional complexity must be explained by laws in its own right. Proponents of nonreductive physicalism are physicalists at the ontological level but nonreductionists at the epistemological level. Hence, physicality in the ontological sense does not imply reduction of psychology to biology, chemistry, or physics. Were one to embark on the epistemological reducibility route, the journey would traverse biology and chemistry and ultimately end in atomic subparticles. Because of emergent properties across levels of complexity, neither the intermediate locales nor the final stop in atomistic physicalism can fully account for human behavior.

As Nagel (1961) explains, there are several necessary conditions for reducibility: They include explicitness of theoretical postulates for each specialized discipline, correspondence or connectability through theoretical terms in common, and derivability from the postulates of the reducing theory. Neither the concepts nor the predicates in psychological theories have representational counterparts in chemistry or physics. Nor do they have an adequate set of bridging principles linking the vocabularies of two theories that are necessary to fulfill the conditions of connectability and derivability. There are lively debates about the required preciseness in linkage between the reduced and reducing theories, whether empirically established links between the two suffice or whether the bridging principles must provide logically necessary conceptual links (van Gulick, 2001).

Consider even the reduction of psychology to biology. Much of psychology is concerned with discovering principles about how to structure environmental conditions to promote given personal and social outcomes and the psychosocial processes through which they produce their effects. This line of theorizing, much of it containing exogenous factors, does not have corresponding concepts in neurobiological theory. How the neuronal machinery works and how to regulate it by psychosocial means are different matters. Each explanatory system is governed by its own set of principles that must be studied in its own right.

For example, knowledge of the locality and brain circuitry subserving learning has little to say about how best to devise conditions of learning in terms of level of intellectual challenge; how to get people to attend to, process and organize relevant information; and whether learning is better achieved independently, cooperatively, or competitively. Psychological science provides a rich body of knowledge regarding the conditions conducive to learning and the psychosocial processes through which they operate. These social determinants reside in the structure of learning environments and in socially rooted incentive systems, enabling opportunity structures, and constraints (Bandura, 1986; Johnson & Johnson, 1985; Rosenholz & Rosenholz, 1981). These determinants operate through modeling, social norms, aspirations, and expectations conveyed in the practices of families, peer relations, school systems, and socioeconomic life conditions (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996, 2001). These are the collective social dynamics of human learning. They have no conceptual counterpart in neurobiological theory and, therefore, are not derivable from it. The optimal learning conditions must be specified by psychological principles. A full explanation of human learning must, therefore, encompass both the psychosocial principles and the subserving neurobiological principles.

System-level emergence calls for theoretical plurality across physical, chemical, biological, psychological, and social structural levels of function with linkage between them rather than reducibility to a single superseding theory. The issue of reductionism in an applied social science must also be evaluated in terms of functional criteria. Can laws at the subatomic or molecular levels tell us how to develop efficacious parents, teachers, executives, or tenacious social reformers? For reasons already given, a psychological level theory is required to provide such guidance.

DISMISSAL OF AGENTIC CONTRIBUTION TO HUMAN FUNCTIONING

The key argument against agentic capabilities is that human action is governed by intricate neural mechanisms that operate outside of one's awareness and control. Hence, thoughts are epiphenomenal events that create an illusion of control but actually have no effect on how one behaves. This is a highly truncated view of how humans exercise control. As explained earlier, people intentionally control functional activities at the macrobehavioral level that activate the subserving events at the microneural level.

Consider some other types of data allegedly demonstrating that conscious cognition has no effect on how one behaves. Rapid activities, such as proficient typing, are said to be much too fast for thought to control the fleet finger movements (Wegner, 2002). Thank goodness it does not have to do so. This line of reasoning fails to distinguish between the differential role that cognition plays during skill development and in its later routinized execution. Human learning is laborious without the aid of instructive thought and awareness of what one is doing. Skill development proceeds through several phases involving a number of different psychomotor functions (Bandura, 1986; Fitts & Posner, 1968). The first phase, in which cognitive factors play an influential role, involves formation of a conception of the skill. The conception serves several proactive functions. It specifies how relevant subskills must be selected, coordinated, and sequenced to suit particular purposes. It also provides the internal standard for translating symbolic conception into proficient performance (Carroll & Bandura, 1990). Without some notion of how the activity is best performed, novices are at a loss as to where to begin, what to do, and what to change. Conceptions are formed on the basis of knowledge gained through social modeling, inferences from response outcomes, and verbal modes of instruction.

These symbolic modes of learning shortcut the toilsome and potentially hazardous form of learning through the rewarding and punishing effects of trial-and-error action. There is considerably more to physical skills, of course, than motor mechanics. Performers have to read multifaceted situations, select effective strategies, anticipate likely outcomes, and improvise performances to suit changing circumstances. Flexible and strategic performance requires a high level of cognitive self-regulation.

Human action is regulated by multilevel systems of control. After proficiency is acquired with cognitive guidance, the skills become routinized and no longer require cognitive control. Their execution is regulated largely by lower level sensory-motor systems in recurrent situations. The actions are run off swiftly without conscious awareness or control. For example, when learning to drive a car with a manual transmission, thoughts about the required operations and the order in which to do them guide the driver's psychomotor learning. After driving becomes a well-integrated routine, people think of other matters while busily driving. Indeed, attending to the mechanics of what one is doing after proficiency is achieved would seriously disrupt skilled performance.

Partial disengagement of thought from proficient action has considerable functional value. Having to think about every skilled action before carrying it out in recurrent situations would consume most of one's precious attentional and cognitive resources and create a monotonously dull inner life. However, when routinized actions fail to produce expected results, cognitive guidance again comes into play. Both the behavior and the changing environmental circumstances are analyzed to identify the source of the problem. New actions are constructed and tested. Control reverts to the lower sensory and motor control systems after an adequate course of action is found and becomes the habitual way of doing things.

Even though thought is disengaged from the mechanics of routinized actions, cognition continues to play an influential role through its strategic function. For example, batters facing a baseball coming at them at 90 miles an hour must anticipate the likely pitch, predict it instantly from subtle pitching cues, and adjust the swing within a split second. They must do their thinking anticipatorily because there is no time for deliberation as the activity is being performed. Coaches amass detailed conditional probabilities of what pitchers are likely to toss in particular situations at particular times with particular batters and relay this information to their batters (Will, 1990). Pitchers are similarly provided with detailed predictive information about the strengths and limitations of the batters they face and what type of pitch to deliver to a particular batter in a particular situation. In the elaborate communications throughout the contest, the athletes try to anticipate and counter each other's strategies at each instant of play. In short, there is a lot of cognitive self-regulation in the contextual orchestration of routinized skills.

It is a common error to equate automation with unconsciousization. The automation of complex skills involves at least three major processes (Bandura, 1986). The first process is *mergerization*, whereby the essential elements of an activity are combined into progressively larger units. When a skill is being learned, the activity is fractionated and some thought must be given as to what to do at each step and transition point in an enactment. Once the routine is put together through the aid of thought and extensive practice, there is no longer any need to think about the subparts and how they should be spatially and temporally coordinated (LaBerge, 1981; Neves & Anderson, 1981). Thought is thus freed for other purposes.

The second process of automation is the establishment of *contextual linkages*. After dealing with the same situation repeatedly, performers eventually learn what works best in that highly predictable situation and respond automatically to predictive situational cues without having to think about what to do. The third process in the automation of skills is a shift in the *locus of attention* from the mechanics of the action to its correlated effects. Actions produce observable effects that indicate what one is doing automatically and suggest needed performance corrections. For example, after driving is routinized, drivers attend to where the car is going. Should it drift off course, they instantly make corrective steering adjustments to get back on track. In shifting their locus of attention with automation, performers monitor the effects of their actions not mechanics of what they are doing.

As the foregoing discussion illustrates, human action is not the product of a unitary process. Most activities contain both cognitively guided and automatic aspects as well as top-down and bottom-up processing. Moreover, the level and form of cognitive guidance changes across phases of skill development and situational circumstances.

Studies by Libet (1985, 1999) are also frequently cited as telling evidence that subconscious neural activity precedes by some milliseconds the conscious intention to act. However, the studies have serious methodological problems that detract from the interpretability of the findings. On the consciousness side, the studies focus on amorphous affective states without discrete onset rather than on an explicit intention to act. Participants were asked to report when they first felt a "desire," "urge," or a "want," whichever was to their liking, and to move their finger or flick their wrist. These are not interchangeable affective states, nor are they intentions as claimed in the articles and citations of the studies. A *want* is a longing; a *desire* is a yearning; an *urge* is an impulse. Intentions represent a determination to act, not a longing or a yearning to do so. Participants were told to adopt a passive mind-set to watch for the appearance of an urge as it emerged "spontaneously" rather than to assert an intention to act. The focus is on self-monitoring of an emerging affective experience. Findings based on onlookers waiting for wants, desires, or urges to rise into view have questionable relevance to a proactive intentional stance.

The alleged "voluntary act" was embedded in a highly constraining and attentionally conflicting context. The activity was consciously prescribed rather than unconsciously decided. Knowing what to do engages second-order control that recruits the subserving lower level neurobiology. Participants had to divide their attention between sensing a desire or urge and an intrusive timing device requiring them simultaneously to fixate on a revolving spot on an analog clock dial, resist following it as it revolved speedily, restrain blinking, note the spot's position when they felt the urge, and commit the judged time to memory for retrospective report. The cognitive processing involved in judging, synchronizing, and remembering clock positions takes some time. The difference between the neural events and the reported affective state is in the milliseconds amidst these multiple conflicting attentional and cognitive demands that create time lags. The generation of an intention centrally, its appearance in awareness, and its temporal registry involve a three-step process with time lags at each step. The actual experience of awareness is undoubtedly much sooner than the recorded time.

Then there is the problem of gauging awareness. Awareness of a conscious event is not a pinpoint experience. Awareness is a progressive event with decisional thresholds of when participants feel sure enough to report a felt desire or urge. Subjective self-report thresholds further increase the latency. The ambiguity of the conscious events being monitored and their timing, multiple conflicting attentional demands, and fuzziness in the precise onset of awareness detract from the interpretability of the temporal ordering of events. If the alleged "intention" is an afterthought, there is the mystery of what sets off the initiatory neural activity. Evidence that preplanning precedes it underscores the need to examine systematically the cognitive activities accompanying the generation of an intention to act. Participants are not sitting idly with a blank mind waiting for an intention to emerge spontaneously. The cognitive activity leading up to a decision to act is part of the instigating condition.

Finally, there is the issue of ecological validity and generalizability. Performance of a purposeless, decontextualized, fractional movement over and over again in multiple sessions may have little to say about the cognitive regulation of action under less fragmented and denuded conditions with wide choice of what to do, often over an extended time course. Continuing with our vacationing motorist, the vacation plan, formulated through a lot of deliberative thought, sets the agenda for a host of preparatory activities and when they should be done. This requires a lot of proactive cognitive regulation. The writings on human agency underscore the influential role played by distal intentions in the cognitive organization and temporal regulation of one's activities (Ismael, 2006; Mele, 1992; Nahmias, 2005). Distal goals structure and give direction to one's activities. But their regulatory influence is best sustained by proximal subgoals that specify what must be done in the here and now to turn a distal vision into reality (Bandura, 1991a). In the case of our vacationer, implementation of the vacation plan does not necessarily require cognitive reminders before performing each preparatory act. In keeping with the dual-level control described earlier, once one knows what needs to be done with commitment to it, calendars, timepieces, and places cue the appropriate actions. For reasons given above, a prescribed isolated finger movement linked to an ambiguous conscious state of equivocal onset is not the type of experimentation on which to rest one's case that thought cannot initiate or regulate action.

A controversial paper by Nisbett and Wilson (1977) is also often cited as evidence that people's actions are governed by unconscious cognitive processes, whereas their conscious cognitions are simply post hoc conjectures about the causes of their actions. A detailed analysis of conceptual and methodological problems raised by their research is presented elsewhere and will not be reviewed here (Bandura, 1986). The present comment centers on the methodological flaw of using retrospective thoughts to confirm that antecedent conscious thoughts have no effect on how one behaves. In the studies, people are asked to explain, after the fact, the reasons for their behavior, or they were presented with preselected factors and asked to judge how they may have influenced their behavior. This retrospective design violates the key temporal criterion that the cause precedes the effect. Tests of whether cognition affects action must assess the thoughts preceding the actions rather than asking participants after undergoing varied experiences to speculate about the causes of their behavior, often with probes that can bias recall by diverting attention from relevant factors or instructing people to judge the influence of irrelevant factors (Adair & Spinner, 1981). Recall can be easily biased, as Loftus (2005) has shown, by what questions are asked and how they are phrased.

In a comprehensive analysis of retrospective thought probes, Ericsson and Simon (1980) explain why the types of probing techniques used by Nisbett and Wilson are ill suited to elucidate the role played by conscious cognition in human behavior. Often the wrong contents are measured by the wrong probes at the wrong time. Ericsson and Simon summarize a great deal of evidence showing that when thoughts are assessed with refined procedures while the activity is ongoing, people verbalize cognitive processes that relate to how they behave.

Another argument against higher level organization and regulation of activities enlists examples of self-organizing systems in which seemingly coordinated activities arise from autonomous subsystems doing their own thing without any overall guidance. The social organization of insect colonies, in which different castes are innately programmed to perform special subfunctions, are often cited. The collective behavior of insects with an inborn repertoire, shaped over millions of years, to execute mechanically a specific action pattern in a particular milieu has little generalizability to the complex functional systems built and operated by humans. To cite but one example, it is a gigantic leap from innately choreographed rituals in an anthill to the extraordinarily innovative and intricately coordinated master plan at NASA to send astronauts to the moon and bring them back safely.

A national vision of space exploration inaugurated this daunting mission. Bringing it to fruition required elaborate central guidance in creating innumerable subsystems and integrating them to function as a complex, interlocking, holistic system. The moon launch required success on the first try rather than mindless trial-and-error groping year after year to evolve a reliable mode of excursion into the inhospitable atmosphere of outer space. The technological evolution relied, in large part, on cognitive ingenuity rather than on morphogenetics. Error elimination was achieved by drawing on specialized knowledge, creating theoretical and process models for computer simulations to test systems under varied possible conditions, and using the results of simulations and the performance of prototypes to redesign and refine the systems. Cognitively guided computational enactments have to supplant, for the most part, physical enactments because of the catastrophic consequences of malfunction in any one of the interlocking subsystems. Once the satellite is airborne, the crew, with the central guidance of their terrestrial overseers, has to solve any unforeseen problems during the mission and the descent. "Houston Control" is not a capricious, epiphenomenal narrative spawned by subpersonal neuronal activity operating autonomously below the level of awareness.

MORAL AGENCY

The exercise of moral agency, rooted in personal standards linked to self-sanctions, is an important feature of an agentic theory of human behavior (Bandura, 1986). In the development of moral agency, individuals adopt standards of right and wrong that serve as guides and deterrents for conduct. In this self-regulatory process, people monitor their conduct and the conditions under which it occurs, judge it in relation to their moral standards and perceived circumstances, and regulate their actions by the consequences they apply to themselves (Bandura, 1991b). They do things that give them satisfaction and a sense of self-worth and refrain from behaving in ways that violate their moral standards because such conduct will bring self-condemnation. Thus, moral agency is exercised through the constraint of negative self-sanctions for conduct that violates one's moral standards and with the support of positive self-sanctions for conduct faithful to personal moral standards.

People have the capability to refrain as well as to act. In the face of situational inducements to behave in inhumane ways, they can choose to resist prepotent social pressures by exerting self-influence. The moral knowledge and standards about how one ought to behave constitute the cognitive foundation of morality. The evaluative self-sanctions serve as the motivators that keep conduct in line with moral standards. Moral thought is translated into moral conduct through this self-reactive regulatory mechanism (Bandura, 1991b).

Moral agents commit themselves to social obligations and righteous causes, consider the moral implications of the choices they face, and accept some measure of responsibility for their actions and the consequences of their actions for others (Keller & Edelstein, 1993). The types of activities that are designated as moral, their relative importance, and the sanctions linked to them are culturally situated. Hence, societies, and even subgroups within them, vary in the types of activities and social practices they consider to be central to morality (Shweder, 2003).

The exercise of moral agency has dual aspects—*inhibitive* and *proactive* (Bandura, 2004a; Rorty, 1993). The inhibitive form is manifested in the power to refrain from behaving inhumanely. The proactive form is expressed in the power to behave humanely. In this dual nature of morality, people do benevolent things as well as refrain from doing harmful things. When individuals strongly invest their self-worth in certain principles and values, they will sacrifice their self-interest and submit to prolonged maltreatment rather than accede to what they regard as unjust or immoral (Bandura, 1999b; Oliner & Oliner, 1988).

Moral standards do not function as unceasing internal regulators of conduct, however. Various psychosocial mechanisms can be used to disengage moral self-sanctions from inhumane conduct (Bandura, 1991b). Selective moral disengagement is most likely to occur under moral predicaments in which detrimental conduct brings valued outcomes. The disengagement may center on sanctifying harmful conduct by moral justification, exonerating social comparison, and sanitizing language. It may focus on obscuring personal agency by diffusion and displacement of responsibility so that perpetrators do not hold themselves accountable of the harm they cause. It may involve minimizing, distorting, or even disputing the harm that flows from detrimental actions. And the disengagement may include dehumanizing, demonizing, and blaming the recipients of the injurious actions. Through selective moral disengagement, people who are considerate and compassionate in other areas of their lives can get themselves to support detrimental social policies, carry out harmful organizational and social practices, and perpetrate large-scale inhumanities (Bandura, 2004b, in press).

In the nonagentic microdeterministic theories reviewed earlier, behavior is the product of nonconscious processes in which environmental inputs activate subpersonal modules that cause the actions. If people's actions are the product of the nonconscious workings of their neuronal machinery, and their conscious states are simply the epiphenomenal outputs of lower level brain processes, it is pointless to hold anyone responsible for the choices they make and what they do. Transgressors should not be held personally accountable for their crimes, police for abusive enforcement practices, prosecutors and jurors for biased sentencing practices, jailers for maltreatment of inmates, and the citizenry for the harmful social conditions that their public policies and practices breed. They can all disclaim responsibility for their actions. Their neural networks made them do it.

Analyses of neuroethics center mainly on the more parochial issues. They include the ethics of pharmacological manipulation of neural systems for self-enhancement and court-ordered management of offenders, the breaching of privacy through functional neuroimaging to detect personal characteristics and cognitive and emotional states, genetic counseling that foretells a disordered future in ways that can be self-fulfilling, and the like (Farah, 2002). The broader moral implications receive little notice, however.

The subpersonal workings of the biological machinery are nonethical. The nonconscious neural processes at the microlevel have neither a sense of personal responsibility nor morality. The issue of morality arises in the purposes to which behavior is put, the means that are used, and the human consequences of the actions. A deterministic thesis that humans have no conscious control over what they do, in fact, represents a position on morality. It is one of moral nonaccountability that is socially consequential. Would a nonagentic conception of human nature erode personal and social ethics that undergird a civil society? How would people create and maintain a civil society if its members are absolved of any personal accountability for their actions?

The incompatibility of nonethical neuronal mechanisms producing ethical and socially responsible conduct poses a formidable challenge for nonagentic theories of human behavior. The proposed solutions usually provide a selective allowance for conscious regulation in the moral domain. In this way one can have automatonization with moral accountability. Libet (1999) voiced concern over the automaton view of human nature and the characterization of humans as blissful illusionists. He proposed a dual-control system in which individuals do not control the initiation of a voluntary act but they can consciously control whether to enact it or veto it. Hence, people can be held responsible for their conduct. Libet's critics vetoed his conscious control function with the regress argument that the conscious veto function is itself the product of preceding unconscious neural processes. Hence, individuals should not be held accountable for what they do not consciously control. This view strips unreservedly any capacity for self-regulative influence over behavior that is morally consequential.

Wegner (2002) also proposed a selective controllability for the pesky morality problem in terms of his conceptual model that Nahmias (2002) calls "modular epiphenomenalism." In this view, environmental inputs activate subconsciously both the neural mechanisms that cause the action and a specialized interpretive module structurally disjoined from the action production system. This complete structural disconnect is puzzling given the intricate neuronal interconnections in the brain (Nahmias, 2002). The interpretive module creates the illusion that one caused the action. This epiphenomenal sense of personal authorship is said to have no causal influence. So what good is an epiphenomenal "self-portrait" if it is merely a "loose end" that cannot affect how one behaves? To get ethicalness into this unconscious subpersonal system, the illusory self-view that one caused the detrimental conduct is invested with casual properties that "can have influences galore" (Wegner, 2004, p. 36). It makes one feel guilty, prompts restitutive acts, and gets one to behave responsibly on future occasions. Rather than resolve the moral incompatibility, the noncausal modularism creates, through unspecified processes, an anomalous epiphenomenon with behavioral causative power, which epiphenomena are not suppose to have. It is also peculiarly selective in its determinative function. The illusory feeling of personal control allegedly affects ethical behavior but can have no effect on other classes of behavior. If the causative power of the epiphenomenal self-view applies broadly, illusory agency begets actual agency for all types of activities. So, for the epiphenomenalists and eliminativists, it is back to the conceptual drawing board on how to make a conscious automaton accountable for its conduct.

Roskies (2006) reassures readers that they need not fear that neuroscience will undermine people's view of themselves as responsible agents. This is because people's judgments of responsibility are unaffected by whether they subscribe to a deterministic or indeterministic view of the world. Given the profusion of interacting neurons, whether a neuron will fire and the type of action potential it generates is probabilistic rather than deterministically inevitable. Hence, Roskies contends that neuroscience cannot undermine freedom and moral responsibility because, at the present state of knowledge, it cannot tell us whether the brain is a deterministic machine. Whether the variability reflects indeterministic processes or complex deterministic ones has to be resolved by physical theory rather than at the level of neurons. For these reasons, in Roskies' view, the ostensible moral problem is the perception of a problem, which she regards as misguided. Neuroethicists and metaphysicalists are not the only ones who have addressed the ethical implications of a neuroscientfic view of human nature. Some folks on the outside have also weighed in with thought-provoking perspectives on this matter (Wolf, 1996).

Morality is not just a matter of perception, however. Simply believing in responsibility is neither personally nor socially consequential unless people have agentic capabilities not only to regulate their conduct, but to create social systems for managing their affairs with authorized rights and the power to implement societally prescribed sanctions. Moral conduct is regulated by three types of sanctions: legal sanctions, social sanctions, and self-sanctions (Bandura, 1986). Their effects are mediated through cognitive processes regarding anticipated risks and potential consequences.

Whether or not a neuroscientific view will erode moral responsibility depends on the form the theorizing takes and the types of experimentation it spawns. In a stimulus driven, bottom-up view in which human behavior is regulated by neuronal processes outside one's awareness and control with thoughts as functionless by-products, as epiphenomenalists and eliminativists contend, it is pointless to hold people responsible for what is beyond their control. If the neuroscientific view recognizes second-order control of brain processes, and the regulative influence of top-down deliberative conscious thought, people can be held accountable for what they do. The latter is a proactive deliberative model the former is a nonconscious reactive one.

The capacity for moral agency is founded on a sense of personal identity, moral standards, and behavioral regulation through self-sanctions (Bandura, 1991b). This ability is acquirable. Social judgments of detrimental conduct are made in terms of personal controllability of the actions. For example, it is within individuals' capacity to stop at a red signal light. A driver who caused a fatal injury by running a red light would be held accountable for his actions. In moral agency, individuals can exercise some measure of control over how situations influence them and how they shape their situations. In the triadic interplay of intrapersonal, behavioral, and environmental events, individuals insert personal influence into the cycle of causation by their choices and actions. Because they play a part in the course of events, they are at least partially accountable for their contribution to those happenings.

Research conducted within the agentic perspective has furthered our understanding of the determinants and processes governing the development and exercise of moral agency (Bandura, 1991b, 1999b). These are rooted in reflexive self-representation and self-reaction. The diverse lines of research clarify how individuals construct moral standards that give meaning and value from the mix of social modeling, the moral values conveyed by evaluative social sanctions to one's conduct and by tuitional means. The theory and verified knowledge specify the processes by which people select, weigh, and integrate morally relevant information in making moral judgments. They verify the self-regulatory mechanisms whereby moral judgments are linked to moral conduct through self-sanctions. And elucidate the psychosocial processes through which moral self-sanctions are selectively engaged and disengaged in the management of moral predicaments.

CONCLUDING REMARKS

Nonagentic theories of human behavior bear resemblance to the behaviorism of yore. In the contemporary reincarnation, stimulus inputs build and strengthen associations automatically and unconsciously. Neural networks become the embodiment of the history of reinforcement. Situational priming bears close likeness to activation by conditioned cues invested with eliciting and signaling properties through correlated experiences. Beliefs, goals, expectation, and other cognitive factors are dismissed as explanatory fictions. It would be the height of irony if the heralded cognitive revolution, which dispatched behaviorism with an indecorous burial, ended up resurrecting it from the presumed dead.

The value of a psychological theory is judged not only by the usual criteria of explanatory and predictive power. In the final analysis, its worth is evaluated by its operative power to solve problems and provide reliable guides for effecting personal and social change. What do nonagentic theories have to offer by way of social utility? They are nihilistic regarding people's capacity to affect the course of events in their lives and are heavily dependent on a stimulus driven approach to regulate behavior subterraneously in the host organism. Many decades ago, Skinner (1971a, 1971b) dubbed cognitive events as explanatory fictions, and attributed freedom, dignity, and the creative products of the mind to the work of environmental forces for which individuals mistakenly take credit. Have we come full circle?

AUTHOR NOTE

Some sections of this chapter include revised and elaborated material from Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on*

Psychological Science, 1, 164–80. I wish to thank Anthony Wagner and Alan Wallace for their comments on an earlier draft of this chapter.

REFERENCES

Adair, J. G., & Spinner, B. (1981). Subjects' access to cognitive processes: Demand characteristics and verbal report. *Journal for the Theory of Social Behavior, 11*, 31–52.

Alland, A., Jr. (1972). *The human imperative.* New York: Columbia University.

Anderson, M., Ochsner, K., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S., Glover, G., Gabrieli, J. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*, 232–35.

Austin, J. H. (1978). *Chase, chance, and creativity: The lucky art of novelty.* New York: Columbia University Press.

Ayala, F. (1974). The concept of biological process. In F. Ayala & T. Dobzhausky (Eds.), *Studies in the philosophy of biology: Reductions and related problems* (pp. 339–56). Berkeley: University of California Press.

Baldwin, C., Baldwin, A., Sameroff, A., & Seifer, R. (1989). *The role of family interaction in the prediction of adolescent competence.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Kansas City, MO.

Baltes, M. M. (1996). *The many faces of dependency in old age.* New York: Cambridge University Press.

Bandura, A. (1973). *Aggression: A social learning analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1982). The psychology of chance encounters and life paths. *American Psychologist, 37*, 747–55.

Bandura, A. (1983). Temporal dynamics and decomposition of reciprocal determinism. *Psychological Review, 90*, 166–70.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1991a). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In R. A. Dienstbier (Ed.), *Perspectives on motivation: Nebraska symposium on motivation* (Vol. 38, pp. 69–164). Lincoln: University of Nebraska Press.

Bandura, A. (1991b). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45–103). Hillsdale, NJ: Erlbaum.

Bandura A. (Ed.). (1995). *Self-efficacy in changing societies.* New York: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A. (1998). Exploration of fortuitous determinants of life paths. *Psychological Inquiry, 9*, 95–99.

Bandura, A. (1999a). A social cognitive theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality* (2nd ed., pp. 154–96). New York: Guilford.

Bandura, A. (1999b). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*, 193–209.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science, 9*, 75–78.

Bandura, A. (2004a). Selective exercise of moral agency. In T. A. Thorkildsen & H. J. Walberg (Eds.) *Nurturing morality* (pp. 37–57). Boston: Kluwer Academic.

Bandura, A. (2004b). The role of selective moral disengagement in terrorism and counterterrorism. In F. M. Mogahaddam & A. J. Marsella (Eds). *Understanding terrorism: Psychological roots, consequences and interventions* (pp. 121–50). Washington, DC: American Psychological Association Press.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*, 164–80.

Bandura, A. (In press) Impeding ecologoical sustainability through selective moral disengagement. *The International Journal of Innovation and Sustainable Development*.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development, 67*, 1206–22.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development, 72*, 187–206.

Bandura, A., & Locke, E. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*, 87–99.

Beccio, C., Adenzato, M., & Bara, B. (2005). How the brain understands intention: Different neural circuits identify the componential features of motor and prior intentions. *Consciousness and Cognition, 15*, 64–74.

Brandtstädter, J., & Baltes-Gotz, B. (1990). Personal control over development and quality of life perspectives in adulthood. In P. B. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 197–224). Cambridge: Cambridge University Press.

Bratman, M. E. (1999). *Faces of intention: Selected essays on intention and agency.* New York: Cambridge University Press.

Bunge, M. (1977). Emergence and the mind. *Neuroscience, 2*, 501–9.

Burns, T. R., & Dietz T. (1992). Cultural evolution: Social rule systems, selection and human agency. *International Sociology, 7*, 259–83.

Carlson, R. (1997). *Experienced cognition.* Mahwah, NJ: Erlbaum.

Carlson, R. (2002). Conscious intentions in the control of skilled mental activity. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 191–228). San Diego, CA: Academic Press.

Carroll, W. R., & Bandura, A. (1990). Representational guidance of action production in observational learning: A causal analysis. *Journal of Motor Behavior, 22*, 85–97.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior.* New York: Springer-Verlag.

Dawson, G., Ashman, S., & Carver, L. (2000). The role of early experience in shaping behavioral and brain development and its implications for social policy. *Development and Psychopathology, 12*, 695–712.

Diamond, M. C. (1988). *Enriching Heredity.* New York: The Free Press.
Dobzhansky, T. (1972). Genetics and the diversity of behavior. *American Psychologist, 27*, 523–30.
Dreifus, C. (2005, May 10). How culture pushed us to the top of the food chain. *New York Times*, p. D2.
Elder, G. (1994). Time, human agency, and social change: Perspectives on the life course. *Social Psychology Quarterly, 57*, 4–15.
Elder, G. (1995). Life trajectories in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 46–68). New York: Cambridge University Press.
Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 87*, 215–51.
Farah, M. (2002). Emerging ethical issues in neuroscience. *Nature Neuroscience, 5*, 1223–29.
Fitts, P., & Posner, M. (1968). *Human performance.* Belmont, CA: Brooks/Cole.
Gardner, R., & Heider, K. G. (1969). *Gardens of war: Life and death in the New Guinea stone age.* New York: Random House.
Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Cambridge: Polity Press; Berkeley, CA: University of California Press.
Gould, S. J. (1987). *An urchin in the storm.* New York: Norton.
Green, C. D., & Vervaeke, J. (1996). What kind of explanation, if any, is a connectionist net? In C. W. Tolman, F. Cherry, R. van Hezewijk, & I. Lubek (Eds.), *Problems of theoretical psychology* (pp. 201–8). North York, ON: Captus.
Greenwood, J. D. (1992). Against eliminative materialism: From folk psychology to Völkerpsychologie. *Philosophical Psychology, 5*, 349–67.
Gruman, J. (2006, December). Quantifying people particles. *Good behavior.* Washington, DC: Center for the Advancement of Health, P.I.
Gusnard, D. (2005). Being a self: Considerations from functional imaging. *Consciousness and Cognition, 14*, 679–97.
Harré, R. (1983). *Personal being: A theory for individual psychology.* Oxford: Blackwell.
Harré, R., & Gillet, G. (1994). *The discoursive mind.* Thousand Oaks, CA: Sage.
Heckhausen, J. (1987). Balancing for weaknesses and challenging developmental potential: A longitudinal study of mother-infant dyads in apprenticeship interactions. *Developmental Psychology, 23*, 762–70.
Hochberg, L., Serruya, M., Friehs, G., Mukand, J., Saleh, M., Caplan, A., Branner, A., Chen, D., Penn, R., & Donoghue, J. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature, 442*, 164–71.
Ismael, J. (2006). Saving the baby: Dennett on autobiography, agency, and the self. *Philosophical Psychology, 19*, 345–60.
Ismael, J. (2007). *The situated self.* New York: Oxford University Press.
Johnson, D., & Johnson, R. (1985). Motivational processes in cooperative, competitive, and individualistic learning situations. In C. Ames & R. Ames (Eds.), *Research on motivation in education: Vol. 2. The classroom milieu* (pp. 249–77). New York: Academic.

Kagan, J. (1981). *The second year: The emergence of self-awareness.* Cambridge, MA: Harvard University Press.

Karniol, R. (1989). The role of manual manipulative stages in the infant's acquisition of perceived control over objects. *Developmental Review, 9,* 205–33.

Keller, M., & Edelstein, W. (1993). The development of moral self from childhood to adolescence. In G. G. Noam & T. G. Wren (Eds.), *The moral self* (pp. 310–36). Cambridge, MA: MIT Press.

Kolb, B., & Whishaw, I. (1998). Brain plasticity and behavior. *Annual Review of Psychology,49,* 43–64.

Korsgaard, C. (1989). Personal identity and the unity of agency: A Kantian response to Parfit. *Philosophy & Public Affairs, 18,* 101–32.

Korsgaard, C. (1996). *The sources of normativity.* Cambridge: Cambridge University Press.

LaBerge, D. (1981). Unitization and automaticity in perception. In J. H. Flowers (Ed.), *Nebraska symposium on motivation* (Vol. 28, pp. 53–71). Lincoln: University of Nebraska Press.

Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer.

Lent, L. (1982). The perception of causality in infants. *Perception, 11,* 173–86.

Levy, R. (1969). On getting angry in the Society Islands. In W. Caudill & T.-Y. Lin (Eds.), *Mental health research in Asia and the Pacific* (pp. 358–80). Honolulu, HI: East-West Center Press.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self.* New York: Plenum.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences, 8,* 529–66.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies, 6,* 47–57.

Locke, E.A. (1994). The emperor is naked. *Applied Psychology: An International Review,* 43, 367–70.

Loftus, E. (2005). A 30-year investigation of the malleability of memory. *Learning and Memory, 12,* 361–66.

Lord, R., & Levy, P. (1994). Moving from cognition to action: A control theory perspective. *Applied Psychology: An International Review, 43,* 335–98.

Mandler, J. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99,* 587–604.

McCaul, K. D., & Malott, J. M. (1984). Distraction and coping with pain. *Psychological Bulletin, 95,* 516–33.

Mele, A. (1992). *Springs of action: Understanding intentional behavior.* New York: Oxford University Press.

Merton, R., & Barber, E. (2004). *The travels and adventures of serendipity.* Princeton, NJ: Princeton University Press.

Millar, W. S. (1972). A study of operant conditioning under delayed reinforcement in early infancy. *Monographs of the Society for Research in Child Development, 37* (2, Serial No. 147).

Millar, W., & Schaffer, H. (1972). The influence of spatially displaced feedback on infant operant conditioning. *Journal of Experimental Child Psychology, 14,* 442–53.

Miller, E., & Cohen, J. (2001). An integrative theory of prefrontal cortex. *Annual Review of Neuroscience, 24*, 167–202.

Moerk, E. (1995). Acquisition and transmission of pacifist mentalities in Sweden. *Peace and Conflict: Journal of Peace Psychology, 1*, 291–307.

Muñoz, R., & Mendelson, T. (2005). Toward evidence-based interventions for diverse populations: The San Francisco General Hospital prevention and treatment manuals. *Journal of Consulting and Clinical Psychology, 73*, 790–99.

Nagel, E. (1961). *The structure of science.* New York: Harcourt, Brace and World.

Nahmias, E. (2002). When consciousness matters: A critical review of Daniel Wegner's *The illusion of conscious will. Philosophical Psychology 15*, 527–41.

Nahmias, E. (2005). Agency, authorship, and illusion for consciousness and cognition. *Consciousness and Cognition, 14*, 771–85.

Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory retrieval signal in the primate temporal cortex. *Science, 291*, 661–64.

Neves, D., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale, NJ: Erlbaum.

Nisbett, R., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–59.

Oliner, S., & Oliner, P. M. (1988). *The altruistic personality.* New York: Free Press.

Ozer, E. (1995). The impact of childcare responsibility and self-efficacy on the psychological health of working mothers. *Psychology of Women Quarterly, 19*, 315–36.

Pascal-Leone, N. Dang, L.G. Cohen, J.P. Brasil-Neto, A Caramota, and M. Hallett. (1995). Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills. *Journal of Neurophysiology*, 74, 1037–45.

Papousek, H., & Papousek, M. (1979). Early ontogeny of human social interaction: Its biological roots and social dimensions. In M. von Cranach, K. Foppa, W. LePenies, & D. Ploog (Eds.), *Human ethology: Claims and limits of a newdiscipline* (pp. 456–78). Cambridge, England: Cambridge University Press.

Powers, W. (1973). *Behavior: The control of perception.* Chicago: Aldine.

Rorty, A. (1993). What it takes to be good. In G. Noam, & T. E. Wren (Eds.), *The moral self* (pp. 28–55). Cambridge, MA: MIT Press.

Rosenholtz, S. J., & Rosenholtz, S. H. (1981). Classroom organization and the perception of ability. *Sociology of Education, 54*, 132–40.

Rosenthal, T., & Rosenthal, H. (1985). Clinical stress management. In D. Barlow (Ed.), *Clinical handbook of psychological disorders* (pp. 145–205). New York: Guilford Press.

Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences, 10*, 419–23.

Rottschaefer, W. (1985). Evading conceptual self-annihilation: Some implications of Albert Bandura's theory of the self-system for the status of psychology. *New Ideas in Psychology, 2*, 223–30.

Rottschaefer, W. (1991). Some philosophical implications of Bandura's social cognitive theory of human agency. *American Psychologist, 46*, 153–55.

Schechtman, M. (1997). The brain/body problem. *Philosophical Psychology, 10,* 149–64.

Searle, J. (2003). *Rationality in action.* Cambridge, MA: MIT Press.

Shweder, R. (2003). *Why do men barbeque? Recipes for cultural psychology.* Cambridge, MA: Harvard University Press.

Skinner, B. F. (1971a). *Beyond freedom and dignity.* New York: Knopf.

Skinner, B. F. (1971b). Why I am not a cognitive psychologist. *Behavior and Philosophy, 5,* 1–10.

Sperry, R. (1991). In defense of mentalism and emergent interaction. *The Journal of Mind and Behavior, 12,* 221–45.

Sperry, R. (1993). The impact and promise of the cognitive revolution. *American Psychologist, 48,* 878–85.

Stagner, R. (1981). Training and experiences of some distinguished industrial psychologists. *American Psychologist, 36,* 497–505.

Teasdale, J.D. (1988). Cognitive vulnerability to persistent depression. C*ognition and Emotion,* 2, 247–74.

Tomita, H., Ohbayashi, M., Nakahara, K., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature, 401,* 699–703.

van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem: A philosophic overview. *Journal of Consciousness Studies, 8,* 1–34.

Watson, J. (1979). Perception of contingency as a determinant of social responsiveness. In E. B. Thoman (Ed.), O*rigins of the infant's socialresponsiveness* (Vol. 1, pp. 33–64). New York: Halsted.

Wegner, D. (1989). *White bears and other unwanted thoughts.* New York: Viking Press.

Wegner, D. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wegner, D. (2004). Précis of *The illusion of conscious will. Behavioral and Brain Sciences, 27,* 649–59.

Will, G. (1990). *Men at work.* New York: Macmillan.

Wilson, E. (1998). *Consilience: The unity of knowledge.* New York: Knopf.

Windmann, S. (2005). What you see is never what you get: Dissociating top-down driven biases in perception and memory from bottom-up processes. In A. Columbus (Ed.), *Advances in psychology research* (Vol. 35, pp. 1–27).

Wolfe, T. (1996). Sorry, but your soul just died. *Forbes, 158,* 210.

7 Free Will Is Un-natural

John A. Bargh

The history of social psychology, and especially its subfield of social cognition, is replete with surprising findings of complex judgmental and behavioral phenomena that operate outside of conscious awareness and even intention (Wegner & Bargh, 1998). Yet the surprising nature of these findings comes no longer from their relative infrequency, for they have become all too commonplace in the research literature. Instead, the surprise comes from the continuing overarching assumption of the field regarding the primacy of conscious will. Based most likely on our (i.e., research psychologists') own subjective experience as human beings, the early process models of each new phenomenon tend to start with the assumption of a major role played by conscious choice and decisions, intention and awareness, in producing the phenomenon in question. Then further findings start coming in showing that, "surprisingly," much of the phenomenon can be explained without need of stages or steps involving conscious intention or awareness.

In the rest of the natural sciences, especially evolutionary biology and neuroscience, the assumption of conscious primacy is not nearly as prevalent as it is in psychology. Thus one goal for the present chapter is to help bring psychology more in line with the rest of the natural sciences, in which complex and highly intelligent design in living things is not assumed to be driven by conscious, intentional processes on the part of the plant or animal, but instead by "blind" natural selection processes (see Dawkins, 1976; Dennett, 1995). As Dennett (1991, p. 251) put it, "in biology, we have learned to resist the temptation to explain *design in organisms* by positing a single great Intelligence that does all the work.... We must build up the same resistance to the temptation to explain *action* as arising from the imperatives of an internal action-orderer who does too much of the work."

The present analysis of the scientific status of the free will concept is therefore in the spirit of the Integrated Causal Model of Tooby and Cosmides (1992) and their forerunners (e.g., Campbell, 1974; Keil, 1979), which take as their starting point not subjective human phenomenology (the strong subjective feeling we have of free will) or the (misleading) model of a serial computer waiting for external commands before operating, but our own deep evolutionary past, in which adaptive features are "designed" into us incrementally by the same processes of natural selection as operated on all other living things. Seen from this different perspective, the recent barrage of "surprising" and sometimes even controversial findings in social psychology regarding consciousness and free will should become less so, and perhaps even predictable.

There are all sorts of confusions regarding the concept of free will, so before proceeding any further, we need to make clear what the psychological concept of free will *is* in order to avoid some of the more common misunderstandings. One of these is between the psychological and the political concepts of free will, so we will start by distinguishing these.

THE PSYCHOLOGICAL VERSUS THE POLITICAL CONCEPT OF FREE WILL

Although the two are often confused, free will as a psychological concept is not the same as free will, or freedom, as a political- or societal-level concept. For one thing, freedom at the political level does not have the same concerns as freedom of the individual will. The main concern of the private, individual domain is with life and survival—the focus is on the home, family, and children. But the public, political domain cannot be mainly concerned with individual survival because its focus on the state of the world transcends the births and deaths of individual members of the society (Arendt, 2005, p. 44). The ancient Greek philosophers, for example, gave a great deal of attention to political freedom, but never even mentioned free will in any of their works. It was St. Paul who discovered the notion of the individual will (Romans, chapter 7), which was then developed more fully by another early Christian thinker, St. Augustine.

Paul wrote that he knew what the good was, he wanted to do it, but could not always do it. In saying this, he introduced the key notion of individual control and responsibility for doing the right thing, and "strength of will" as an important determinant of whether an individual will successfully do it instead of yielding to the temptations of the flesh. St. Augustine built upon this idea in his explanation for how evil could coexist in a world along with a wise, good, all-powerful God. For Augustine, the just reward of eternal heaven and individual salvation required the possibility within each individual of both good and bad behavior (see Neiman, 2002). Accountability at the final judgment, in the divine scheme of things, required free individual control over actions.

Thus free will as a psychological concept concerns the individual, and actions that are under the individual's power to perform. However, free will or freedom as a political- or societal-level concept (which is often confused with the psychological sense of the notion), concerns the plurality and actions that depend on or are restricted by the consent or cooperation of others. In other words, a person has free will in the political sense if he or she is not acting under coercion, if his or her actions are not determined by external forces; as long as the causes of the action are internal to the individual (motives, needs, preferences, desires), the person is acting "freely."

In this light, a commonly heard argument in favor of free will—that it exists "because I can choose to (or not to) do *X*," makes sense from the political perspective, but not within the psychological domain, where it only begs the question. The psychological status of freedom of the will depends on whether those *choices* are determined or not by identifiable forces—and, as we shall see, on whether such choices are even necessary in the first place. This is, of course, a higher standard than the litmus test for free will in politics: The latter requires only an absence of external determination; the former requires an absence of both external and internal determination of the action.

As Arendt (1978, p. 448) concluded, "the Christian and modern notion of free will has no ground in political experience." The philosopher of science Karl Popper (1965, p. 128) has noted that phenomena that are apparent at the level of resolution used in one branch of science (in this case, political science) may disappear at the finer levels which are the basis for other branches (here, psychology): "The 'dots' produced by the coarse 'screens' used in the reproductions of photographs in our daily papers are satisfactory representations when looked at superficially, but cannot stand closer inspection with a magnifying glass. So, too, the reproductions of the world by our forms of intuition and categories break down as soon as they are required to give a somewhat closer representation of their objects, as is the case in wave mechanics and nuclear physics."

INFLUENCED VERSUS *DETERMINED*

The psychological issue of whether free will exists thus boils down to whether *undetermined choices* of action exist and occur. No one today would deny that people have preferences, motivations, desires, goals, and so on, and that these at least *influence* what we do. This is after all the very subject matter of psychological science. But the doctrine of free will within psychology holds as axiomatic (see Locke & Kristof, 1996) that the *choices* made on the basis of these influences are free, made by a consciousness that is the source of "original intentionality" (Searle, 1983). Now we have distilled the essence of the question of free will, in the psychological domain: Are behaviors, judgments, and other higher mental processes the product of free conscious choices, as *influenced* by internal psychological states (motives, preferences, etc.), or are those higher mental processes *determined* by those states? The *influence* model can be likened to an executive officer who takes suggestions from subordinates as to what to do but nonetheless makes the decisions; the *determination* model has those subordinates directly in charge with no need of an independent Decider.

Yet any scientific—as opposed to philosophic—approach to the question of free will cannot rely upon extraphysical explanatory concepts, as Searle (1983) did with the concept of original intentionality, and as John Locke did before him with his mind-first cosmology (see Dennett, 1991). Locke had argued that *mind* was the originator of thought and action, but that nothing (save one's own past personal experience) caused mind. Similarly, for Searle, only humans (not other living things) are said to have original intentionality, by which he meant that intentions (the will) originate in the mind and are not themselves the causal product of any physical or mechanical forces. As Konrad Lorenz (1962, p. 23) admonished us, "it is the duty of the natural scientist to attempt a natural explanation before he contents himself with drawing upon factors extraneous to nature." Treating free will as a force outside the laws of nature in the Locke/Searle manner is similar to how intuition and creativity have long been popularly viewed as being due to some kind of mysterious "spark" or quasi-magical process. In all three cases, the argument that the phenomenon is an originator and not itself caused by some other process is actually just an admission that *we don't know* what causes it; as Spinoza (1677/1951, p. 134) put it, "men believe themselves to be free, simply because they are conscious of their actions, and unconscious of the causes whereby those actions are determined."

THE PHENOMENAL PAST VERSUS THE PHENOMENAL (NEAR) FUTURE

It is not merely ignorance of the underlying causes of our actions that gives us such a strong subjective sense that they are spontaneous and thereby "free," however. There is also a fundamental difference in *time perspective* between our experiences of our own behavior versus that of other people (Arendt, 1978). Our experience of the outside world of others' actions is retrospective in nature, it deals with what they have done. Scientists analyzing the reasons for behavior deal with *faits accompli*, behaviors that exist in the past, which is the world of causation. But our own internal perspective on our own behavior is *prospective* in its focus, dealing with the relatively uncertain and unpredictable future. The philosopher Harry Frankfurt (1971) has noted that we tend to invoke the notions of choice or free will only when describing our *own* behavior, *not* that of other people. And more recently, Pronin and Kugler (in press) have documented this attributional difference experimentally; choice or deliberation does not come up in accounts of why others did what they did, only for one's own behaviors. When we are accounting for other people's behavior we are like scientists, because the perceived and experienced behavior of others is in the past; but only we as individuals have privileged access to our own phenomenal state prior to acting. Arendt (1978) makes the following comments:

> In the perspective of memory, that is, looked at retrospectively, a freely performed act loses its air of contingency under the impact of now being an accomplished fact, of having become part and parcel of the reality in which we live. The impact of reality is overwhelming to the point that we are unable to "think it away"; the act appears to us now in the guise of necessity. . . .

Once things have happened, and have receded into the past, they become part of the world of facts, of causes, and we just naturally, even inescapably feel that they were determined, caused, and that nothing else could have happened. We may not be able to predict what will happen but once it does, we feel we "knew it all along," and believe what happened was inevitable. This fundamental difference between our subjective certainty and confidence about the past, versus our uncertainty and trepidation about the future, manifests itself in many judgment biases that have been documented by decision researchers. The *hindsight bias* (e.g., Fischhoff, 1975; Hawkins & Hastie, 1990) is our feeling after an event has occurred that "we knew it all along" coupled with an actual inability to recall what it was we had expected or predicted prior to the outcome. The *just-world bias* (Lerner, 1980) is our tendency to believe that things are as they are because that is how they ought to be. The *status quo effect* is related to the just-world bias, in that people are biased to prefer the current state of affairs (the status quo) and are reluctant to change it (e.g., Samuelson & Zeckhauser, 1988; Kay, Jimenez, & Jost, 2002). (A standard experimental demonstration of the status quo effect is to tell one group of participants that a certain policy is in place and an alternative has been proposed, and switching the in-place and alternative policies for another group: When asked which policy they prefer, both groups prefer the one that they had been told is already in place, even though they are quite different.) The past for us feels phenomenally determined and, after the fact of course, is experienced as having been inevitable.

But we are not very good at predicting the future; even experts such as in sports and politics are often wrong in their prognostications and still, after the fact, freely discuss what has just happened in causal terms as if it were inevitable and "of course" this is what happened. In the *hindsight effect,* our very memories are biased in the direction of having predicted the outcome that actually occurred (see Ross, 1989), and as Dawes (1993) has noted, over time this creates our strong belief that the world is more predictable than it really is. In actuality, the world is full of randomness and uncertainty. The novelist Milan Kundera uses this as a running theme of his novels, relentlessly pointing out to the reader the many coincidences and chance happenings that had to have occurred for his protagonists to have ever met in the first place (e.g., *The Unbearable Lightness of Being*). Dawes (1993) has shown mathematically that it has to be the case that retrospective analysis of the causes of an event—say, the clinical psychologist's locating the cause of a patient's depression in her mother's cold and distant attitude toward her as a child—significantly overestimates the actual strength of the causal relation, compared to purely predictive studies of the same relationship. He concludes the following from his analysis:

> What are the implications?...We are interested in predicting the unusual from the unusual. When we do so, however, a fundamental asymmetry results. The degree of predictability appears to be systematically greater when the analysis is retrospective than when it is prospective....Even those who believe in destiny or God's Plan for the future do not claim to know exactly what it is. In contrast, "history" appears to be understandable, whether it is our own or that of others we retrospectively assess....We tend to derogate the role of random influences in how we got to here, in contrast to their role in where we will go from here. For example, past personal sufferings or economic recessions are easily explained in terms of psychological or economic conditions or "forces" ("causes"), while whether we are entering such an unhappy period now is a matter of "speculation." (Dawes, 1993, pp. 7, 17)

Especially in contrast with our subjective sense of the determined past, the experience of our own behavior in the present seems particularly spontaneous and "free." Because we do not experience at the same time all of the unconscious influences and impulses that produced that behavior (see Bargh & Morsella, 2007), our phenomenal experience is hugely biased in the direction of feeling that we have much more freedom than we actually do. But our feelings, like much else about us, have evolved because of their adaptive significance and are essential for normal cognitive functioning (e.g., Gray, Schaefer, Braver, & Most, 2005; Tranel, Bechara, & Damasio, 2000). We have also learned that feelings of being in control are far more beneficial to our functioning than are feelings of helplessness; thus these subjective feelings of free will are one of the "positive illusions" (Taylor, 1989) we hold dear. Yet this benefit is irrelevant to the scientific status or truth value regarding the actual existence of free will; however positive and adaptive the feeling, it is still an illusion.

CONSEQUENCES OF THE UNPREDICTABLE FUTURE FOR THE EXISTENCE OF FREE WILL

The fact of the uncertain and unpredictable future has strong implications for the existence and scope of free will. Imagine that your job was to design and create a device that had to function far into the future, long after you were no longer around, or a space probe that would eventually get too far away from Earth to receive your commands. In these examples, the success and survival of your creation will eventually depend on its making its own decisions, based on local conditions you cannot anticipate, but decisions nonetheless derived from the general purposes and parameters you originally designed into it.

In his seminal work *The Selfish Gene*, Dawkins (1976) drew just such an analogy between how genes "design" (through the blind process of natural selection) their "survival machines" on which they depend for their propagation into future generations. Because natural selection processes, through gene mutations, operate over vast units of time, they cannot in any way adapt in real time to changes or events in the environment. Thus, genetic controls over behavior are relatively inflexible and can't adapt quickly to sudden changes in the environment. (This is largely why 99% of the species that ever existed are now extinct.) All they can do is to instantiate the few specific principles most likely to be adaptive even far into the future—such as strong motives to survive, to eat, to reproduce—along with those general principles or strategies that give the organism some adaptive advantage that increases the gene's chances of being passed down to the next generation. In harmony with Dawes (1993) and other decision scientists as to our overestimation of the predictability of the world, Dawkins (1976, p. 55) notes that "prediction in a complex world is a chancy business. Every decision that a survival machine takes is a gamble, and it is the business of genes to program brains in advance so that on average they take decisions that pay off."

LEVELS OF CONTROL: GENETIC, CULTURAL, PSYCHOLOGICAL

It is for this reason that evolution has shaped us to be *open-ended* systems (Mayr, 1976, p. 695). This gives room for "fine-tuning" of the human infant to local conditions, as through culture and learning. The genetic determinants of our behavior reflect only the most basic truths that are important for our (and the genes that we carry; Dawkins, 1976) survival and reproduction, truths that have been abstracted out of eons of our ancestors' evolutionary history. The mechanism through which genes drive our present-day behavior is through evolved motives (Tomasello, Carpenter, Call, Behne, & Moll, 2005). The active goal or motive is the "local agent" by which the genetic influence finds expression (see also Neuberg, Kenrick, Maner, & Schaller, 2004). As Tomasello et al. (2005) and others have noted, this is how evolution works—through motives and strategies, desired goals and end-states, that we seek to get to from whatever starting point in history and geographical location the cards of fate have dealt us—*not* through rigid and fixed responses to specific events or stimuli, because these cannot be anticipated; the world itself is evolving and changing with the turn of the centuries and millennia.

Take, for example, the pied flycatcher, an English bird that navigates by night based on the patterns of stars in the sky (Grocott, 2003). Clearly this is an evolved skill, but it cannot be entirely innate, and must be an open-ended adaptation, because the pattern of stars and constellations in the sky is constantly changing as the Earth moves on its own path through the galaxy (as are the other celestial bodies in the night sky). The night sky in the northern (and southern, of course) hemisphere is very different today than it was just a few hundred years ago! So it is not that these birds have evolved to have a perfect map of a constant night sky available to them, as an internal representation, to guide their flight. What they have evolved is this: the general ability to navigate in very precise fashion from the current pattern of stars and constellations, the pattern that exists when they are born. Their parents take them out several times for night flights after they are born and able to fly, and it is during these practice flights that they absorb—*download*, if you will—the current, contemporary pattern of stars. Evolution has, of course, also given them the ability to do this downloading, and to represent it, and to base their flight behaviors on it—but this is the *hardware*. For the hardware to work, the little birds need to download the *software* of the current star pattern; thus it is an open-ended ability or system, one that must adapt to current, local conditions if it is to work at all.

And this is what human culture gives us, in analogous fashion: the local conditions, mainly social, of the world (and the particular region of the world) into which we happen to have been born. Dawkins (1976) noted that phenotypic plasticity (the openness of the evolved system) enables the infant to absorb, entirely automatically and unconsciously, "an already invented and largely debugged system of habits in the partly unstructured brain" (p. 193). This cultural knowledge is a giant step toward adaptation to the current local environment that the genetic determinants of our development could in no way accomplish. In this way, a human infant can be relocated immediately after birth to any place and any culture in the world and that child will adapt to and speak the language of that culture just as well as any child born there. As Dennett (1991, p. 200) pointed out, "one of the first major steps a human brain takes in the massive process of postnatal self-design is to get itself adjusted to the local conditions that matter the most: it swiftly (in 2 or 3 years) turns itself into a Swahili or Japanese or English brain." Thus one's particular language and culture are not genetically determined; it is only the ability to become, so quickly, a member of any culture that is determined by one's genes.

Culture, including language, norms, values, and so on, is "downloaded" after birth, and it reduces greatly the unpredictability of the child's world, and his or her uncertainty as to how to act and behave in it. As the cultural anthropologist Dan Sperber (1980, p. 26) has argued, "public representations come before private ones, a child is born into a world full of public representations and is bombarded with them from the first moments of life." Again, the culture we soak up in such an amazingly fast and thorough manner (as in the case of language acquisition) early in life—as well as continually thereafter—exerts powerful constraining and controlling influences on our choices and behavior in life. These are as strong as evolutionary forces, if not stronger, for people are willing to die for their culture, their country, their religion, which cannot possibly be what their selfish genes had in mind for them.

There is yet a third level of adaptation, producing an even finer level of predictability and control for the emergent adult human. This is *learning,* the psychological level of adaptation, in which the child's particular experiences shape him or her with expectations of what happens next—given event A, event B is what usually follows—and this knowledge of outcomes helps to direct and constrain the child's behavior at a finer level of local adaptation than even the general culture.

Thus, *evolution* gives us the general motives and strategies for survival, *culture* gives us the general rules and knowledge of how to live in the particular part of the world and the particular group of people into which we happen to have been born, and *learning* from our own direct experience gives us even finer-grained understanding and predictive anticipations. Note, however, that these are not *independent* influences; as Dawkins (1976, p. 193) points out, our ability to absorb culture depends on phenotypic plasticity (the openness of the evolved system). This in turn depends on genetic variation—that is, we as humans *acquired* the ability to acquire culture through natural selection. Similarly, in the case of learning, for it to be adaptive we must be predisposed (through natural selection processes) to learn about only certain aspects of the environment over others, because of the overwhelming amount and variety of information that constantly impinges upon us (Lorenz, 1962; see also Campbell, 1960; Norretranders, 1998; Plotkin & Odling-Smee, 1982).

PREFERENCES

We are also predisposed, from evolution but also culture and early learning, to prefer certain objects and aspects of our environment over others. We are often guided by our "feelings," intuitions, and "gut reactions"; indeed these preferences have been shown to be indispensable to adaptive behavior, especially to prioritizing what is important versus not so important to do or attend to (e.g., Damasio, 1996; also Schwarz & Clore, 1996). These feelings or guides do not arise out of thin air: Our present preferences are derived from those that served adaptive and functional ends in the past. As Dennett (1995) argued, the perspective of a conscious observer is a more sophisticated descendant of the ancient perspectives of the first replicators, who simply divided their worlds into good versus bad. The psychologist and evolutionary epistemologist Donald Campbell (1974) called these "shortcut processes" because they save us from having to figure out, each of us individually from scratch, what are the good and helpful things and which are the dangerous and unhelpful things.

In Campbell's (1974) view, knowledge processes are part of a nested hierarchical system. A basic tenet of evolutionary theory is that evolution builds gradually on what it has to work with at that moment; changes are gradual and incremental. Knowledge gained at a lower level of blind selection, the shortcuts and other "good tricks" (Dennett, 1995) that consistently worked over our long-term evolutionary past, can be fed upward as a starting point—appearing as a priori knowledge, the source of which we are unaware. These are the bases of our mysterious hunches and intuitions, and even our creative new answers and solutions were given a starting boost by our evolved tendencies.

That our consciously expressed preferences are based on such primordial preferences, at least as a starting point, helps to make some recent surprising findings more sensical. There is a long-standing research domain in social psychology on *automatic attitudes* (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986; also Bargh, Chaiken, Govender, & Pratto, 1992), in which a person's attitudes are shown to become active automatically upon the mere presence of the attitude object (or its name) in the stimulus environment (see Ferguson, 2007, for review). In other words, pretty much everything we encounter is evaluated, unintentionally and unconsciously, as either good or bad immediately after we encounter it (i.e., within 250 milliseconds). But this research had always focused on actual attitude objects with which people had some degree of experience (e.g., tuna, poison, birthdays, flies), so the understanding was that this effect required a consciously formed attitude or preference to occur. Duckworth, Bargh, Garcia, and Chaiken (2002), however, showed that the same effect occurred, within the same paradigm, for *novel* attitude objects that were nonrepresentational in nature, such as snippets of abstract art. This was surprising because the participants had never encountered these stimuli before, yet they immediately classified them as good or as bad within a quarter of a second.

Interestingly, we had selected these novel attitude objects on the basis of pretesting with another group of participants, who were exposed to them consciously and then took as long as they needed to say whether they liked them or not. We then took the ones most of these pretest participants liked, and the ones most of them disliked, and used them in the experiment with a new set of participants. And for these experimental participants, the "good" and the "bad" novel attitude objects showed the same automatic attitude effect, and of nearly identical size, as the real attitude objects used in previous research.

Now for this effect to be obtained, the experimental participants had to have evaluated the novel attitude objects that they had never seen before and to which they were exposed for only 250 milliseconds in the same way as had participants who had as much time as they needed and who evaluated the objects consciously and intentionally. Because logically the pretest participants who evaluated the items consciously would be expected to also have the same automatic evaluative processes as the experimental participants, one can conclude that they too had "known" the goodness versus badness of the stimuli immediately, and this must have served the as the basis for their longer, deliberative conscious evaluation. Again, from the starting assumption that automatic evaluative processes are based only or mainly on consciously made evaluations based on personal experience with them (e.g., Fazio et al., 1986), the Duckworth et al. (2002) findings of the automatic evaluation of entirely novel objects are surprising and even controversial. If, however, unconscious preference processes serve as the starting point on which our subsequent conscious evaluative processes then build, as Campbell (1974) argued, those findings make perfect sense.

Moreover, research has shown these automatically activated attitudes and preferences to be directly connected to behavioral mechanisms (just as are priming effects from the behavior of other people around us, the perceived goals of those people, and so on). Chen and Bargh (1999) showed that participants are faster to make approach movements of the arm (pulling a lever toward oneself) when responding to positive attitude objects, and faster to make avoidance movements (pushing the lever away) when responding to negative attitude objects, even though their conscious task in the experiment was just a reaction time task to "knock off the screen" the names of these objects when they appeared. And this unconscious behavioral tendency to approach what is good and avoid what is bad extends even to novel objects never encountered before; Duckworth et al. (2002) replicated the Chen-Bargh automatic evaluation-action link for the novel attitude objects in their study.

Immediate, unconsciously produced evaluations can produce even more powerful and abstract behavioral effects than simple arm movements. In a recent study by Todorov et al. (2005), ratings of competence of U.S. congressional election candidates, based solely on facial appearance with the faces presented for just 1 second each, predicted the outcomes of the 2004 U.S. congressional elections better than chance—for example, 68.8% of the Senate races in 2004 were successfully predicted from these immediate, intuitive inferences. Voting choices, of course, are important decisions and widely assumed to be based on deliberate, conscious, and rational processes, yet these findings suggest that even important decisions are influenced and predicted by immediate unconscious evaluative processes.

Here is another example that most people find surprising, again because it involves important life decisions. It has long been known that we have a strong preference and liking for people who are similar to ourselves in appearance, attitudes, and beliefs, and this plays a significant role in interpersonal attraction (Byrne, 1971). Recent research has shown that this similarity-liking effect extends to new people who resemble significant others such as our parents (Andersen & Chen, 2002), although people are not aware of and do not report any such resemblance as a factor in their liking. The similarity effect is so strong, in fact, that it extends even to preferences for places to live and occupational choices that are similar to ourselves in merely superficial ways.

For example, compared to what you'd expect by chance alone, there are more people named Ken who moved to live in Kentucky, Florences who moved to Florida, and more named Louis who moved to St. Louis; there are more Dennises and Denises who become dentists and Lauras and Lawrences who become lawyers, compared to people with names that do not share letters with these occupations. If your first or last name begins with "H," you are more likely than chance to own a hardware store, and if one of your names begin with "R," you are more likely to own a roofing company, with "C" a computer company, and with "T" a travel business (for many such studies, see Jones, Pelham, Carvallo, & Mirenberg, 2004; Pelham, Mirenberg, & Jones, 2002). This is not at all to say that name-letter similarity is the only basis for our choice of domiciles and professions, but that it is a statistically significant influence on those choices. Most people find this, well, surprising, and it is clearly an unconscious influence as no one would claim name-letter overlap as a reason for making these important life choices.

Other superficial similarities, such as sharing a birthday with another person, produce the same powerful effects. Walton and Cohen (2006) manipulated whether their participants had the same birthday or not with a fellow student, described in a (fictitious) newspaper article (which listed the student's birth date incidentally in a brief biography sideline to the main article) as having just won a prestigious award for mathematical achievement. Compared to the different-birthday condition, those students who shared a birthday with the award winner actually had higher grades at the end of the semester in their math classes.

What is the basis for such "implicit egotism"? Evolutionary biologists have traced this tendency to a *kin selection bias* that gave our genes a further reproductive and survival advantage (Williams, 1966); this is one of the pillars of "selfish gene" theory (Dawkins, 1976), in which genes, not we, as individuals, are argued to be the unit of natural selection. According to selfish-gene theory, over evolutionary history we tend to like those who resemble us because they tended to share genes with us—resemblance in appearance was correlated with resemblance in genetic makeup. (Note how this extends to one's parents as well as in Andersen's "transference" research, because they each share half of our genes.) And the contemporary social psychological research described above is showing just how powerful this similarity-preference effect is, as it extends to all sorts of features stored within our self-representations (note again how it is the *mental representation* that mediates these effects), beyond just physical appearance, to our names, our particular birth dates, and other self-attributes important to our identities.

CONTEXTUAL PRIMING: NATURAL, UNCONSCIOUS ADAPTATION TO THE PRESENT

The open-ended nature of our evolved design has also caused us to be highly sensitive and reactive to the on-line, right-now *present*. Social psychologists over the past 30 years have been studying these sensitivities under the rubric of *contextual priming effects*. The unpredictability of the future, as it relentlessly approaches us, requires us to be continuously *reactive* to unfolding events. Because we can't know with any degree of certainty what will happen in advance (in most natural situations), we have to react and adapt to what is currently going on—and the evolved design of our minds causes the on-line presence of these events and objects and people to automatically activate our internal representations of them. With the activation of the representations comes, concomitantly, all of the internal information (affect, goals, behaviors, knowledge) relevant to responding back to the current environment (Bargh, 1997).

Events in the current situation automatically activate or *prime* their corresponding mental representations inside of us, and this passive activation persists for a short time thereafter (Higgins, 1996; Higgins, Bargh, & Lombardi, 1985). This is important because the increase, over baseline levels, in activation makes these concepts more *accessible* to further activation by events, creating what Bruner (1957) termed *perceptual readinesses* to interpret the events around us (particularly social behaviors, which nearly always can be interpreted in multiple ways) in line with what has just recently happened. Over time, accessibilities can become chronic, reflecting the long-term probabilities in one's environment. If the two forms, temporary and chronic, happen to conflict with other—such as when your typically quiet and bookish Uncle Albert starts to dance on top of the dinner table after Thanksgiving dinner—the temporarily accessible construct ("wild and crazy guy") overrides the chronically active construct (Higgins & Bargh, 1987) for the time being. The chronic expectations do soon return to preeminence when the temporary accessibility effect decays back to baseline (Higgins, Bargh, & Lombardi, 1985; Bargh, Lombardi, & Higgins, 1988).

That the temporary accessibility effect trumps the chronic one when the two are in conflict is another example of the adaptive nature of accessibility effects, because for temporary priming effects to be useful at all, they have to win out in the short term. Otherwise, we'd always be at the mercy of our long-term past, and never able to adapt to changing circumstances, except only after considerable (probably painful, being so wrong about things for so long) additional experience, sufficient to change the chronic set. In these ways, priming and construct accessibilities tune us to the long- and short-term probabilities within our current environment and represent another way in which unconscious processes are open-ended and flexibly adapt to current conditions.

Imitation and Mimicry

The priming effects of people's behavior and other situational features on us extend beyond influencing our perceptual interpretations and expectations, however. They also directly influence our own behavior, beginning soon after birth. Infants naturally learn much about how to behave by mere passive imitation of fellow children and also their adult caretakers; indeed, Meltzoff (2002) concluded from decades of researching this phenomenon that infants can imitate body movements and facial acts at birth, and that this ability represents a "primordial connection between infant and caretaker" (p. 19).

These imitative impulses, triggered by the perceived behavior of others, continue to be activated throughout one's life, causing children and adults to have default tendencies to act the same as those around us are acting—producing behavioral and emotional contagion effects. Thus, how other people are acting around us in the present is yet a further unconscious influence or guide as to how we ourselves should act. As Dawkins (1976) pointed out, the best behavioral strategy (from the point of view of evolution and adaptation) "depends on what the majority of the population is doing" (p. 69; see Maynard Smith, 1982; Maynard Smith & Parker, 1976). Thus, "blindly" or unconsciously adopting what others around you are doing, especially in new situations or with strangers, makes good adaptive sense as a default option or starting point for your own behavior. This tendency, and its unconscious and unintentional nature, has been repeatedly demonstrated in human adults in the research of Chartrand and colleagues (e.g., Chartrand & Bargh, 1999; Chartrand, Maddux, & Lakin, 2005; Lakin & Chartrand, 2003). People don't know and even don't believe after you tell them that they had engaged in these imitative behaviors—on several occasions they insisted on seeing their own videotapes before they would believe it. Not only do people tend to adopt the physical behavior (posture, facial gestures, arm and hand movements) of strangers with whom they interact, without intending to or being aware they are doing so, such unconscious imitation also tends to increase liking and bonding between the individuals—serving as a kind of natural "social glue" (Chartrand & Bargh, 1999; Giles, Coupland, & Coupland, 1991; Lakin & Chartrand, 2003).

Importantly, Meltzoff has shown via experiments with infants given pacifiers that their imitative abilities, even soon after birth, depend on an *internal-representation* of the outside behavior. Although the infants are prevented from immediate imitation by the pacifier in their mouths, when it is taken out they do then imitate the prior facial gesture of the caretaker. Starting from birth, then, these internal representations are the interface between environment and mind, and the basis of unconscious contextual priming influences.

It is by reference to these same internal representations, then, that the adult human being is wide open to external influences, and even control, over his or her behavior. Fifty years ago, B. F. Skinner (1957) attempted to show that all behavior was under the direct control of the stimulus environment, but as single reflex acts, without reference to any internal mental representations. The transparent failure of this attempt was one reason for the cognitive revolution in psychology (Chomsky, 1959; Koestler, 1967; Neisser, 1967). However, by theoretically extending the reach of external stimuli to the internal representations of the environment that they automatically activate (e.g., types of behavior, goals, social groups, specific other people), much of what Skinner (1957) claimed in terms of direct environmental control over the higher mental processes has now been validated in contemporary research on priming effects across a variety of psychological phenomena (see Bargh & Ferguson, 2000). Yes, the internal mechanisms are cognitive, they are "mental," but they are not dependent on a homuncular "ghost in the machine" (Ryle, 1949) as they can operate entirely unconsciously.

MIND READING AND GOAL PRIMING

Along with the physical aspects of others' behavior, we are also highly attuned to the intentions or purposes that underlie that behavior. Meltzoff (1995) and Tomasello et al. (2005) have shown in studies of 18-month-old children that they encode and represent not only the actual behavior of their caretakers but what the caretaker is trying to do as well—if, for example, the caretaker is repeatedly unsuccessful at putting some toys away in a box, the child will not imitate the unsuccessful attempts but the intended act of putting the toy away properly. With adult participants, Knuf, Aschersleben, and Prinz (2001) deployed some sleight of hand with mirrors in order to disentangle perceptual from intentional mimicry (priming) and showed that perception-behavior effects are governed more strongly by representations of intended events than by representations of perceived events. Prinz (2002) concluded from these studies that "we have every reason to believe that intentional induction is no less automatic than perceptual induction . . . [and] that understanding actions and their consequences in terms of their underlying intentional semantics develops very early in life and then remains so deeply rooted in our cognitive machinery that we have no way to escape from it" (p. 159).

Such "mind-reading" ability has now been well documented in the developmental literature on "theory of mind" and has found strong recent support in cognitive neuroscience as well, with the discovery of *mirror neurons* in the premotor cortex that become active both when you perceive a given type of action and when you engage in that action yourself (see Bargh, 2005; Frith & Wolpert, 2003; Meltzoff & Prinz, 2002). This tight, automatic connection between our perceptual and our actional representations suggests that we are prewired to have behavioral and goal-pursuit tendencies in line with those around us, whose behavior we are currently perceiving and decoding in terms of underlying intent.

This is the evolutionary structural support for yet another type of priming effect that has been demonstrated in social cognition research over the past decade: goal priming. Information-processing goals such as judgment and memorization, achievement goals such as high performance on a task, and interpersonal goals such as competition or cooperation have all been successfully primed using subtle manipulations (sometimes even subliminally)—that is, activated unconsciously, without the person's awareness or intent. Across many such studies, the consistent outcome is that unconscious goal pursuit produces the same outcomes as when that goal is pursued consciously and intentionally (see Bargh, 2005; Chartrand & Bargh, 2002; Dijksterhuis, Chartrand, & Aarts, 2007). For example, subliminal priming of the goal of cooperation causes participants playing the role of a fishing company to voluntarily put more fish back into a lake in order to replenish the fish population, compared to a control condition (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Troetschel, 2001).

Moreover, the qualities of the underlying process seem to be the same, for participants with interrupted nonconscious goals tend to want to resume and complete a boring task even when they have more attractive alternatives, and will show more persistence on a task in the face of obstacles, compared to control conditions (Bargh et al., 2001)—features long known to hold for conscious goal pursuits (Lewin, 1926). People will even be in a worse mood after "failing" at a primed goal they didn't even know they had, and a happier mood after "succeeding" at that goal (Chartrand & Bargh, 2002), just as in conscious goal pursuits. And in none of these studies are participants able to accurately report on what they'd just done, in terms of the goal they'd been pursuing. For example, the correlation between self-reported cooperation and actual cooperative behavior during the experiment was around .30 in a condition in which participants were explicitly instructed to cooperate, but near zero in the cooperation-priming (unconscious) condition, though the two conditions produced equivalent amounts of cooperation (Bargh et al., 2001, Study 2).

Note in regard to this finding of unconscious motivation to cooperate that Tomasello et al. (2005) have identified cooperation and helping as an evolved motive, one that they argue is the key difference between humans and other primates. Tetlock (2002) has similarly argued, with supportive evidence, for evolved social motives of accountability to the others in one's social group (the intuitive politician mind-set) and of enforcing group standards on others (the prosecutorial mind-set). This is important regarding the present argument against the existence of free will (at the psychological level), because many would take from the nonexistence of free will that people have no responsibility for their actions and therefore can act entirely selfishly and without regard to the consequences of their actions for others. Thus even if behavior is (multiply) determined and "free will" does not exist at a psychological level, part of the determination of behavior includes motivations to be responsible to others and to be vigilant about and act against their own potential irresponsibilities.

IF NOT FROM FREE WILL, WHERE DO OUR BEHAVIORS ORIGINATE?

Historically, free will has been the answer to the question of where our actions originate, of where they come from in the first place. But there is now an alternative answer to the question because (as outlined above) there is no shortage of ideas or suggestions from our unconscious as to what to do in any given situation. There are a multitude of adaptive behavioral impulses generated at any given time from our evolved motives and preferences, from our cultural norms and values, from our own past experiences in that situation, as well as from what other people are currently doing in that situation. And all of these unconscious inputs from the past and present of our world—even activated attitudes and preferences—have been shown to be directly connected to behavioral mechanisms, that is, to *action tendencies.* Recent neuroscience research has confirmed the close, automatic connection between our perceptual and our actional representations in both primates and humans. Similarly, priming research in social cognition has documented how sensitive we are to the behavior and goals of those around us and how we find it a positive, rewarding experience to be doing the same thing as they are, and work on automatic attitudes has shown our immediately generated likes and dislikes to be directly connected to muscular action tendencies to either approach or avoid the object.

Before there was consciousness, there already were all the unconscious modules and components that evolved to serve adaptive ends—selective sensitivity to the important and dangerous aspects of the environment (in large part, other members of our own species with whom we directly competed for the same needed resources from the environment), basic motivations to survive, eat, reproduce, to avoid what was known to be bad for us and to approach that which was good for us. According to Dennett (1991), primate brains are "based on millennia of earlier nervous systems; they were regularly flooded with multimodal information, and this gave them a new problem, one of higher-level control. There wasn't a convenient captain already on board, so these conflicts had to sort themselves out without any higher executive" (p. 188). Conflict exists at every level—for example, in sexual reproduction, the male and female alleles can be "in conflict" with each other (e.g., blue vs. brown eyes)—but for the most part, these conflicts are resolved for us unconsciously (see Morsella, 2005).

CONNECTING A PARALLEL MIND TO A SERIAL WORLD

This means that the "blind" unconscious mental modules that serve us so well from a functional perspective must be capable of some form of adaptive integration to produce single, serial decisions and behavior in real time. (This is the problem of connecting the parallel brain to the serial world in which we can only do one thing at a time; Bargh, 1997, 2006.) This supposition will again help us make sense of two more current puzzles, one that was posed 50 years ago, and another just recently.

The first of these is how quickly and suddenly, in terms of evolutionary timescales, we acquired language (see Pinker, 1994). It was not gradual, and did not depend on our brains growing to a certain critical size, for Neanderthal brains, which did not have language, were if anything larger than our contemporary brains (Calvin, 1989). Language is a complex skill that could not possibly be acquired so quickly in young children through normal, slow, trial-and-error learning processes (Chomsky, 1959); it develops spontaneously in nearly all children worldwide regardless of their levels of intelligence. The language production mechanism, "through use of 'phrase structure,' takes a web of thoughts and outputs them in form of words spoken one at a time, without a conscious effort or formal instruction, and is deployed without awareness of its underlying logic" (Pinker, 1994, pp. 101–2).

The speed with which we acquired language as a species, and the exponential advances in culture and knowledge we've made since then (see Diamond, 1992), suggests that as an ability it piggybacked or was "scaffolded" onto an existing structure, or what Dennett (1995) called a "good trick"—a solution nature has come up with for a problem that tends to be used over and over again in nature (e.g., the independent evolution of eyes in many different species). The evolutionary theorist Calvin (1989) argued similarly that innate language abilities themselves are quite recent, even rushed, additions to our genetic makeup, and as such are very likely exaptations of previously existing sequencing circuitry in the brain. What this means for present purposes is that not only did sophisticated unconscious modules evolve that give us today the building blocks of adaptive motives, preferences, and behavioral impulses, all operating unconsciously, there was also evolved a mechanism to integrate or interface these separate, parallel inputs into serial behavioral and judgmental responses. Our ability to take a vague thought and have it come out of our mouths in a complete coherent sentence, the production of which happens unconsciously, is a paramount example of this. *It is not something we need consciousness or free will for.* It is not the case that notwithstanding all of these wonderful adaptive unconscious inputs, we still need a central conscious executive, operating spontaneously and freely, to make behavioral decisions based on these inputs. All of those separate types of input, as documented above, have their own direct connections to behavioral mechanisms. And there also must have been some mechanism to integrate the multiple parallel unconscious inputs into serial responses, because this is a problem we faced as a species in the distant past before the development of consciousness and language, as evidenced by the opportunistic exploitation or "co-option" of the mechanism by language.

Note, too, in this regard the recent "surprising" findings of Dijksterhuis, Bos, Nordgren, and van Baaren (2005), who showed that better decisions are made when a person is distracted while making them than when able to devote total conscious attention and deliberation to the process. In these studies, across a variety of domains, participants were first presented with the relevant facts, and then made decisions as to the best house to buy or which soccer team would win a particular match. Some then had to do a secondary task and so were distracted during the decision time interval, and others were not. Those who were distracted consistently outperformed the "conscious" group in these choices (the studies were designed so that there was a best or optimal solution by objective standards).

From the perspective of modern decision theory these are very surprising findings, but from the present perspective it makes sense that left to its own devices of integrating various disparate pieces of information and coming to the best answer—the task for which the unconscious mind evolved for eons prior to the late add-on of conscious processing—the unconscious route worked best. The conscious process, unlike the unconscious, was driven by whatever particular heuristic or theory the participant might possess as to how to make the decision (e.g., "Eindhoven never loses on Sunday," "Stone houses have the best resale value"), and other research has shown such lay theories to be of dubious validity or value (see Nisbett & Wilson, 1977; Wilson & Brekke, 1994); moreover, conscious but not unconscious processes suffer from capacity limitations on the processing of multiple pieces of information in parallel (e.g., Miller, 1956). Again, the finding that people make better decisions when the integration processes are unconscious, rather than conscious, is surprising and controversial only if one is assuming that free will or conscious processes are required to solve such integration and combination problems.

THE PRIMACY OF THE UNCONSCIOUS

Multiple sources of behavioral impulses, coming from evolution, culture, learning, and the current environment, reverse the usual assumptions of how we generate our behavior. As we stand, each moment, on the threshold of the near future, experiencing the uncertainty and spontaneity of our own actions, we are not aware of and do not experience all of these unconscious influences acting upon us. Our subjective phenomenology has given us the strong sense, difficult to overcome, that our ethereal free will is the source of our behaviors, judgments, and goal pursuits.

Again, given as well the field of psychology's meta-assumption of the primacy of conscious will, the extensive documentation of unconscious controls from our distant and recent past and our present seem surprising and controversial. But reversing the causal assumption and recognizing the substantial role played by unconscious forces of evolutionary design, cultural assimilation in early childhood, and our minds as wide open to environmental priming influences, makes these and other similar findings much less controversial and more understandable. The lines of priming research described above show how action and motivational tendencies can be put into motion and cause us to behave in a certain way, without our being aware of the source of those tendencies.

But there are other demonstrations outside of priming research. Take, for example, Libet's (1986) well-known *time of intention* studies, Wegner's (2002) *misattribution of will* studies, and Baumeister, Bratslavsky, Muraven, and Tice's (1998; Muraven, Tice, & Baumeister, 1998) striking demonstrations of *ego depletion* effects. In the Libet paradigm, participants are free to make a button-pressing or other response whenever they choose (simulating the state of free will) and are asked only to note when (by reference to a sweep-hand clock in front of them) they had made the intention to respond. Libet at the same time was measuring brain activation potentials associated with the instigation of action (i.e., the P300 wave). The "surprising" finding was that the action potential consistently came 300–500 milliseconds *before* the participant's conscious awareness of intending to make the response. Consistent with the present argument that our action impulses are generated for us through unconscious mechanisms, the impulses, even in this paradigm emphasizing free will or action, came prior to the person's conscious awareness of having made them (but see Dennett & Kinsbourne, 1992).

Wegner's research (Wegner & Wheatley, 1999) makes this point in a different way, by showing how people's feeling of having willed a given event to occur is an *attribution* or inference (not a direct readout of actual causation) based on key variables such as the timing of their thoughts of performing the action relative to the action occurring, through a novel paradigm in which these variables could be manipulated without the person's knowledge. The right combination of these variables produced feelings in the participants of having willed the event when in fact it was not under their control at all. Finally, the ego depletion studies (Baumeister et al., 1998), in which making even a simple conscious choice or decision significantly decreased a person's ability to engage in self-control, were surprising because they showed how little conscious self-regulatory capacity we actually have (far less than one would expect if conscious choice and intention was required for nearly all of our mundane behaviors, as long argued by Bandura [1986] among others).

Each of these lines of research converge on the conclusion that the will is not the source of ideas of what to do next. Impulses come from a variety of unconscious sources, as described above, prior to consciousness becoming aware of them (Gazzaniga, 1985). Twenty years ago, Libet's (1986) claim seemed preposterous to many, because we did not know much about where these mysterious "impulses" came from (if not from conscious will), but now we do know enough about them to bring Libet's claim into the mainstream.

CONCLUSIONS

I have argued here for a new way of looking at the issue of free will, one that begins with the assumption of mainly unconscious instead of conscious causation of action and phenomenal experience, and that is better aligned with our knowledge of the rest of nature, in which examples of amazing, complex yet unconsciously operating design (in animals and plants) are plentiful (see Dawkins, 1976; Dennett, 1995). As has often been noted (e.g., Blackmore, 1999; Dawkins, 1976, p. 67), the value of a new perspective can be seen in terms of what phenomena it can readily explain that were previously difficult to account for. Among such phenomena that were surprising from the starting assumption of conscious choice and free will, but which make sense within the present perspective of the primacy of unconscious forces, are (1) the automatic evaluation of novel objects, (2) the immediate connection between automatic evaluation and behavioral (motoric) tendencies, (3) the name-letter and birth-date effect on important life decisions, (4) the unconscious mimicry of others' behavior, (5) unconscious goal pursuit over time in the absence of ability to accurately self-report on one's intentions, (6) the very recent and rapid acquisition of language abilities in evolutionary history, (7) that unconsciously made decisions involving integration of relevant features are superior in quality to consciously made ones, (8) the misattribution of free will, (9), that brain-wave impulses to act precede conscious awareness of the intention to act, and (10) the scarcity of conscious self-regulatory capacity. To me, this is rather impressive evidence for the value of the new perspective, in which unconscious, not conscious, causes are primary, and unconscious, not conscious, processes are assumed at the outset of any new line of inquiry.

Regarding the psychological concept of free will, the evidence reviewed above, along with the substantial banks of knowledge already gained in the other natural sciences, leads to the conclusion that there is no need to posit the existence of free will in order to explain the generation of behavioral impulses, and there is no need to posit free will in order to explain how those (unconscious) impulses are sorted out and integrated to produce human behavior and the other higher mental processes. The phenomenological feeling of free will is very real, just as real for those scientists who argue against its actual existence as for everyone else, but this strong feeling is an illusion, just as much as we experience the sun moving through the sky, when in fact it is we who are doing the moving. Each of us lives in a difficult to predict present and near future, which includes our own behavior in it, and which therefore makes our behavior feel spontaneous and undetermined—but what we don't experience, yet which are just as real, are the multitude of unconscious influences and determinants of what we think, act, and feel.

Finally, as psychologists who are also natural scientists, we need to keep in mind that the "unconscious mind" is the rule in nature, not the exception. It is, perhaps, time for us to stop being so surprised.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by Grant R01-MH60767 from the U.S. Public Health Service. Many thanks to Roy Baumeister, Ap Dijksterhuis, Ezequiel Morsella, and Jim Uleman for feedback on an earlier version.

REFERENCES

Andersen, S. M., & Chen, S. (2002). The relational self: An interpersonal social-cognitive theory. *Psychological Review, 109*, 619–45.

Arendt, H. (1978). *The life of the mind*. New York: Harcourt.

Arendt, H. (2005). *Responsibility and judgment*. New York: Schocken.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.

Bargh, J. A. (2005). Bypassing the will: Towards demystifying behavioral priming effects. In R. Hassin, J. Uleman & J. Bargh (Eds.), *The new unconscious*. Oxford, UK: Oxford University Press.

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36*, 147–68.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*, 893–912.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: The automaticity of higher mental processes. *Psychological Bulletin, 126*, 925–45.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Troetschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81*, 1004–27.

Bargh, J. A., Lombardi, W., & Higgins, E. T. (1988). Automaticity of chronically accessible constructs in Person x Situation effects on person perception: It's just a matter of time. *Journal of Personality and Social Psychology, 55*, 599–605.

Bargh, J. A., & Morsella, E. (2008). The primacy of the unconscious. *Perspectives on Psychological Science.*

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego-depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*, 1252–65.

Blackmore, S. (1999). *The meme machine.* New York: Oxford University Press.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*, 123–52.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.

Calvin, W. H. (1989). *The cerebral symphony: Seashore reflections on the structure of consciousness.* New York: Bantam.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67*, 380–400.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–63). La Salle, IL: Open Court.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893–910.

Chartrand, T. L., & Bargh, J. A. (2002). Nonconscious motivations: Their activation, operation, and consequences. In A. Tesser, D. Stapel, & J. Wood (Eds.), *Self and motivation: Emerging psychological perspectives* (pp. 13–41). Washington, DC: American Psychological Association Press.

Chartrand, T. L., Maddux, W., & Lakin, J. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. Hassin, J. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 334–61). New York: Oxford University Press.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*, 215–24.

Chomsky, N. (1959). Review of *Verbal Behavior* by B. F. Skinner. *Language, 35*, 26–58.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London B, 351*, 1413–20.

Dawes, R. M. (1993). Prediction of the future versus an understanding of the past: A basic asymmetry. *American Journal of Psychology, 106*, 1–24.

Dawkins, R. (1976). *The selfish gene.* New York: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York: Simon & Schuster.

Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer. *Behavioral and Brain Sciences, 15*, 183–247.

Diamond, J. (1992). *The third chimpanzee: The evolution and future of the human animal*. New York: HarperCollins.

Dijksterhuis, A., Aarts, H., & Chartrand, T. L. (2007). Automatic behavior. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes.* Philadelphia: Psychology Press.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006, February 17). On making the right choice: The deliberation-without-attention effect. *Science, 311*, 1005–7.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 6*, 515–19.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–38.

Ferguson, M. J. (2007). Automatic evaluation. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes.* Philadelphia: Psychology Press.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–99.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 1*, 5–20.

Frith, C., & Wolpert, D. (Ed., 2003). *The neuroscience of social interaction.* New York: Oxford University Press.

Gazzaniga, M. S. (1985). *The social brain.* New York: Basic Books.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied psycholinguistics.* New York: Cambridge University Press.

Gray, J. R., Schaefer, A., Braver, T. S., & Most, S. B. (2005). Affect and the resolution of cognitive control dilemmas. In L. Barrett, P. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 67–94). New York: Guilford.

Grocott, D. F. H. (2003). Maps in mind: How animals get home. *Journal of Navigation, 56*, 1–14.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgment of past events after the outcomes are known. *Psychological Bulletin, 107*, 311–27.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. T. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–68). New York: Guilford.

Higgins, E. T., & Bargh, J. A. (1987). Social perception and social cognition. *Annual Review of Psychology, 38*, 369–425.

Higgins, E. T., Bargh, J. A., & Lombardi, W. (1985). The nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 59–69.

Jones, J.T., Pelham, B.W., Carvallo, M., & Mirenberg, M.C. (2004). How do I love thee?, Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology, 8*, 665–83.

Kay, A., Jimenez, M. C., & Jost, J. T. (2002). Sour grapes, sweet lemons, and the anticipatory rationalization of the status quo. *Personality and Social Psychology Bulletin, 28*, 1300–1312.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective.* Cambridge, MA: Harvard University Press.

Knuf, L., Aschersleben, G., & Prinz, W. (2001). An analysis of ideomotor action. *Journal of Experimental Psychology: General, 130*, 779–98.

Koestler, A. (1967). *The ghost in the machine.* London: Hutchinson.

Lakin, J., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science, 14*, 334–39.

Lerner, M. J. (1980). *The belief in a just world.* New York: Plenum.

Lewin, K. (1926). Vorsatz, wille, und bedürfnis [Intention, will, and need]. *Psychologische Forschung, 7*, 330–85.

Libet, B. (1986). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*, 529–66.

Locke, E. A., & Kristof, A. L. (1996). Volitional choices in the goal achievement process. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 363–84). New York: Guilford.

Lorenz, K. (1962). Kant's doctrine of the a priori in the light of contemporary biology. *General Systems, 7*, 23–35.

Maynard Smith, J. (1982). *Evolution and the theory of games.* New York: Cambridge University Press.

Maynard Smith, J., & Parker, G. A. (1976). The logic of asymmetric contests. *Animal Behaviour, 24*, 159–75.

Mayr, E. (1976). *Evolution and the diversity of life.* Cambridge, MA: Harvard University Press.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–50.

Meltzoff, A. N. (2002). Elements of a developmental theory of imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 19-41). New York: Cambridge University Press.

Meltzoff, A. N., & Prinz, W. (2002). *The imitative mind: Development, evolution, and brain bases.* New York: Cambridge University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psychological Review, 63*, 81–97.

Morsella, E. (2005). The function of phenomenal states: Supramodular interaction theory. *Psychological Review, 112*, 1000–1021.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*, 774–89.

Neiman, S. (2002). *Evil in modern thought: An alternative history of philosophy.* Princeton, NJ: Princeton University Press.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Neuberg, S. L., Kenrick, D. T., Maner, J. K., & Schaller, M. (2004). From evolved motives to everyday mentation: Evolution, goals, and cognition. In J. Forgas & K. Williams (Eds.), *Social motivation: Conscious and unconscious processes* (pp. 133–52). New York: Cambridge University Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–59.

Norretranders, T. (1998). *The user illusion.* New York: Viking.

Pelham, B. W., Mirenberg, M. C., & Jones, J. K. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology, 82*, 469–87.

Pinker, S. (1994). *The language instinct.* New York: William Morrow.
Plotkin, H. C., & Odling-Smee, F. J. (1982). Learning in the context of a hierarchy of knowledge gaining processes. In H. C. Plotkin (Ed.), *Learning, development, and culture* (pp. 443–71). New York: John Wiley & Sons.
Popper, K. R. (1965). *Conjectures and refutations: The growth of scientific knowledge* (2nd ed.). London: Routledge and Kegan Paul.
Prinz, W. (2002). Experimental approaches to imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 143–62). New York: Cambridge University Press.
Pronin, E., & Kugler, M. B (in press). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology.*
Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96,* 341–57.
Ryle, G. (1949). *The concept of mind.* Chicago: University of Chicago Press.
Samuelson, W., & Zeckhauser, R. J. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7–59.
Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 433–65). New York: Guilford.
Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind.* New York: Cambridge University Press.
Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.
Sperber, D. (1980). The epidemiology of beliefs. In C. Fraser & G Gaskell (Eds.), *The social psychological study of widespread beliefs* (pp 25–44). Oxford: Clarendon Press.
Spinoza, B. de (1951). *Ethics* (proposition III, part II). In R. H. M. Elwes (Ed. & Trans.), *Spinoza: The chief works* (Vol. 2). New York: Dover. (Original work published 1677)
Taylor, S. E. (1989). *Positive illusions.* New York: Basic Books.
Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109,* 451–71.
Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005, June 10). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623–26.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28,* 675–91.
Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.
Tranel, D., Bechara, A., & Damasio A. R. (2000). Decision-making and the somatic marker hypothesis. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 1047–61). Cambridge, MA: MIT Press.
Walton, G., & Cohen, G. (2006). *Mere belonging.* Manuscript submitted for publication, University of Waterloo.
Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed.). Boston: McGraw-Hill.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*, 480–92.

Williams, G. C. (1966). *Adaptation and natural selection: A critique of current evolutionary theory*. Princeton, NJ: Princeton University Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117–42.

8 The Automaticity Juggernaut—or, Are We Automatons After All?

John F. Kihlstrom

Writing at the very beginnings of scientific psychology, James (1890/1980) considered the "automaton-theory" proffered by La Mettrie (1748/1749), that humans were conscious automatons, but automatons nonetheless—thus erasing Descartes's distinction between man and animals. Although admitting that "habit covers a very large part of life" (p. 109), James concluded that "to urge the automaton-theory upon us, as it is now urged, on purely *a priori* and *quasi-metaphysical* grounds, is an *unwarrantable impertinence in the present state of psychology*" (p. 141, emphasis original).

James was skeptical about the automaton-theory, but he implied that he was open to new scientific evidence that might prove the theory to be true after all. According to some psychologists, that moment has arrived. In 1999, the *American Psychologist* published a special issue entitled "Behavior—It's Involuntary" describing the concept of automaticity as a "fundamental breakthrough in the understanding of motivations, free will, and behavioral control"—the breakthrough being the recognition that, in the ordinary course of everyday living, we really are operating on automatic pilot. As the guest editor wrote, "We perceive ourselves to have far more control over our everyday behavior than we actually do.…[T]he source of behavioral control comes not from active awareness but from…mental activations of which we are unaware and environmental cues to which we are not consciously attending that have a profound effect on our behavior (Park, 1999, p. 461).

Whereas once psychologists might have believed that unconscious and automatic processes were exceptional, if not downright pathological, apparently we have now come to understand that they are the normative rule and conscious control the exception. How did this happen, and why?

THE ROOTS OF AUTOMATICITY

The modern notion of automaticity has its origins in research on attention (Kahneman, 1973), and was particularly inspired by the familiar Stroop effect induced by discrepancies between the identities of color words and the ink in which they are printed (MacLeod, 1991). Apparently, subjects cannot help processing the meaning of the word, and this interferes with the manifest task of naming the color in which the word is printed. Studies of reading and visual search also contributed to the elaboration of the concept (LaBerge & Samuels, 1974; Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977, 1984). By the end of the 1970s, cognitive psychologists had settled on a distinction between automatic (or effortless) and controlled (or effortful) information processing in which automatic processes were defined in terms of four core properties:

1. *Inevitable evocation:* Automatic processes are inevitably engaged by the appearance of specific environmental stimuli, regardless of the person's conscious intentions, deployment of attention, or mental set.
2. *Incorrigible completion:* Once evoked, they run to completion in a ballistic fashion, regardless of the person's attempt to control them.
3. *Efficient execution:* Automatic processes are effortless, in that they consume no attentional resources.
4. *Parallel processing:* Automatic processes do not interfere with, and are not subject to interference by, other ongoing processes—except when they compete with these processes for input or output channels, as in the Stroop effect.

Hasher and Zacks (1979) offered an expanded concept of automaticity, in which automatic processes are not only independent of any intentions, but also independent of any other personal or environmental conditions, such as emotional states, distractions, or stresses. Moreover, provided that the person is neurologically intact, they suggested that performance on automatic tasks would be uncorrelated with individual differences in arousal level or intelligence, or with group differences associated with race, ethnicity, socioeconomic status, or education. These features proposals are consistent with a general principle of *automatic invariance*, but it is not clear that these features should have the same status as the canonical attributes outlined above—which Bargh (1994) aptly dubbed "the four horsemen of automaticity."

The contemporary concept of automaticity has its roots in biological and ethological studies of reflexes, taxes, and instincts—all of which are conceived as involuntary responses to eliciting or releasing stimuli—and in traditional analyses of classical and instrumental conditioning, in which environmental stimuli are held to acquire the power to elicit specific responses. Thus, although some automatic processes may be innately specified, others appear to become automatized through extensive practice. Because they consume no attentional resources, automatic processes leave no consciously accessible traces in memory. In fact, widespread acceptance of the notion of automaticity was an early milestone in the revival of interest in unconscious mental life (Kihlstrom, 1987; see also Hassin, Uleman, & Bargh, 2005). In theory, at least, automatic processes are unconscious in the strict sense of the term: They are neither available to conscious awareness nor amenable to conscious control.

FROM COGNITIVE TO SOCIAL PSYCHOLOGY—AND BEYOND

The concept of automaticity was an important advance in cognitive theory, as it offered a resolution of the dispute between early- and late-selection theories of attention (Pashler, 1998). According to the early-selection view, preattentive, preconscious processing was limited to analyses of the physical features of a stimulus; in theory, analysis of meaning required the conscious deployment of attention. According to the late-selection view, even meaning analyses were conducted preattentively. Automaticity theory permitted complex, semantic analyses to be carried out preattentively, and thus preconsciously, so long as they were automatized—for example, through extensive practice. In later developments, automaticity became detached from attention theory, and was reinterpreted in terms of memory (J. R. Anderson, 1992; G. D. Logan, 1988). In addition, cognitive psychologists began to develop experimental paradigms, such as the process-dissociation procedure (L. L. Jacoby, 1991), by which they could estimate the contributions of automatic and controlled processes to task performance.

Following its embrace by cognitive psychology, the concept of automaticity quickly spread to other domains, particularly personality and social psychology. For example, Nisbett and Wilson (1977) clearly had automaticity in mind when they argued that we are consciously aware of the *contents* of our minds, such as beliefs and attitudes, but unaware of the *processes* that generated those contents: "We may have no direct access to higher order mental processes such as those involved in evaluation, judgment, problem solving, and the initiation of behavior" (p. 232).

Similarly, Langer asserted that most social interactions are unreflective and mindless, following highly learned, habitual scripts that require very little conscious attention and deliberation: "[M]indlessness may indeed be the most common mode of social interaction" (E. Langer, Blank, & Chanowitz, 1978,

p. 641). "Unless forced to engage in conscious thought, one prefers the mode of interacting with one's environment in a state of relative mindlessness. . . . This may be the case, because thinking is effortful and often just not necessary" (E. J. Langer, 1978, p. 40).

Along these lines, Taylor and Fiske (1978) argued that people are "cognitive misers" laboring under limited cognitive capacity, and preferring "top of the head" judgments to reasoned, thoughtful appraisals. Smith and Miller (1978) were perhaps the first to explicitly invoke the concept of automaticity, as it was then emerging in cognitive psychology, in a commentary on the Nisbett/Wilson paper. From their point of view, limitations on introspective access occurred because salient social stimuli are processed, and responded to, automatically.

Thereafter, a number of social psychologists explicitly referred to the concept of automaticity in designing and interpreting experiments on attitudes and social judgments. For example, Higgins and King (1981) distinguished between two sources of automatic priming effects on social judgments: chronic and temporary. Bargh (1982) showed that presentation of self-relevant adjectives over the unattended channel in a dichotic listening task could disrupt shadowing performance, after the matter of the "cocktail-party phenomenon"; and that parafoveal presentation of hostile trait adjectives could bias interpretation of the "Donald story" used in studies of impression formation and person memory (Bargh & Pietromonaco, 1982).

By the end of the 1980s, the concept of automaticity had been applied across a large number of domains in personality and social psychology, including prejudice, the self-concept, emotion, trait ascriptions, and ruminative thought. A landmark volume edited by Uleman and Bargh (1989) contained chapters detailing the role of automatic, unintended thoughts in a variety of domains, including the activation of self-beliefs and ruminations in anxiety and depression; the influence of feelings on thought and behavior; the ascription of personality traits and the formation of characterological impressions; heuristic information processing in persuasion; and ironic rebound effects.

THE AUTOMATICITY JUGGERNAUT GAINS MOMENTUM

After 1989, the concept of automaticity proliferated rapidly through personality and social psychology (Bargh, 1994). A PsycINFO search reveals that prior to 1975, the terms *automatic* or *automaticity* had appeared in the abstracts of only 29 articles published in personality and social psychology journals—and most of these had to do with automatic writing and other aspects of spiritualism. Another 6 were added by 1980; in the 1980s, there were 40 such articles; and in the 1990s, 115 (for comprehensive coverage of these studies, see

D. M. Wegner & Bargh, 1998). As this chapter was being proofread, in late 2007, the new millennium had added 240 new papers—a geometric increase of interest in automaticity, as opposed to the almost perfectly linear increase in the total number of articles published over the same span of time (see also Bargh, 2007).

In 1997, the *Journal of Experimental Social Psychology* devoted an entire issue to the role of automatic processes in stereotyping and prejudice; the *Journal of Personality and Social Psychology* followed suit in 2001. In between, Greenwald, Banaji, and their colleagues introduced the Implicit Association Test, intended to reveal hidden prejudices, and based on the assumption of an automatic association between certain social stimuli and judgments of "good" or "bad" (Greenwald, McGhee, & Schwartz, 1998; Nosek, Greenwald, & Banaji, 2005),

Of course, the concept of automaticity gained popularity in its home territory of cognitive psychology, as well—but with a difference. Cognitive psychologists have maintained a distinction between automatic and controlled processes, and have spent a great deal of effort in assessing their differential contributions to task performance—as in the process dissociation paradigm (e.g., L. L. Jacoby, 1991). At first, social psychologists followed suit, resulting in a number of *dual-process* theories of attitudes, persuasion, and the like, which described the interplay between automatic and controlled processes (e.g., Chaiken & Trope, 1999). Fairly quickly, however, this balanced perspective began to be replaced by a more single-minded focus on automaticity. For example, Gilbert (1989, p. 189) argued for the benefits of "thinking lightly about others." And Bargh (2000, p. 938) argued that even intentionally controlled behavior was ultimately automatic in nature, "controlled and determined" by "automatically operating processes." Thus, rather than taking a balanced view of the differential roles of automatic and controlled processing in social interaction, some social psychologists seem to have embraced a view of social thought and action as almost exclusively automatic in nature.

This evolutionary development can be clearly seen in the work of John Bargh, who has been one of the foremost proponents of the concept of automaticity within social psychology. In 1984, writing on "the limits of automaticity," Bargh was critical of Langer's position that social interaction proceeded mindlessly:

> A better summary of the mindlessness studies would be that . . . when people exert little conscious effort in examining their environment they are at the mercy of automatically-produced [*sic*] interpretations. . . . Automatic effects are . . . typically limited to the perceptual stage of processing. There is no evidence . . . that social behavior is often, or even sometimes, automatically determined. (Bargh, 1984, pp. 35–36)

But only 5 years, later, his position had shifted considerably, as in the editorial introduction to *Unintended Thought:*

> As most social psychological models implicitly assumed the role of deliberate, calculated, conscious, and intentional thought, the degree to which unintended [automatic] thought did occur in naturalistic social settings became of critical importance.... Langer (1978) emphatically rejected the assumption of deliberate, conscious, thought as typically underlying social behavior.... Our own research programs have followed in this tradition.... (Bargh & Uleman, 1989, pp. xiv–xv)

And in his own contribution to that volume, he writes,

> Is this to say that one is usually not in control of one's own judgments and behavior? If by "control" over responses is meant the *ability* to override preconsciously suggested choices, then the answer is that one *can* exert such control in most cases.... But if by "control" is meant the actual *exercise* of that ability, then the question remains open.... My own hunch is that control over automatic processes is not usually exercised.... [I]t would appear that only the illusion of full control is possible, as the actual formation of a judgment or decision.... A fitting metaphor for the influence of automatic input on judgment, decisions, and behavior is that of the ambitious royal advisor upon whom a relatively weak king relies heavily for wisdom and guidance. (pp. 39–40)

Only one year later, Bargh took a further step, asserting that automaticity pervades the information processing system, such that automatically evoked mental representations automatically generate corresponding motives, which in turn automatically generate corresponding behaviors (Bargh, 1990; Bargh & Gollwitzer, 1994). Thus, merely reading words related to rudeness or politeness can affect whether a subject will interrupt the experimenter's conversation, whereas reading words related to the elderly stereotype will lead subjects to walk more slowly down the hall (Bargh, Chen, & Burrows, 1996; see also Ferguson & Bargh, 2004).

In a chapter describing "the automaticity of everyday life," Bargh continued to expand the role of automatic processes:

> [T]he more we know about the situational causes of psychological phenomena, the less need we have for postulating internal conscious mediating processes to explain these phenomena.... [I]t is hard to escape the forecast that as knowledge progresses regarding psychological phenomena, there will be less of a role played by free will or conscious choice in accounting for them.... That trend has already begun..., and it can do nothing but continue. (Bargh, 1997a, p. 1)

Later in the same chapter, Bargh asked, "Is consciousness riding into the sunset?": "Automaticity pervades everyday life, playing an important role in creating the psychological situation from which subjective experience and subsequent conscious and intentional processes originate..." (p. 50).

Actually, in the typical Western, the hero rides into the sunset only after rescuing the sheriff, vanquishing the villain, and kissing the girl—a pretty good situation. The image Bargh really seems to have in mind is of the sun *setting on* consciousness—or, perhaps, consciousness on an ice floe, like the elderly Eskimo, floating out to sea. But just in case the reader missed the message, Bargh quickly repeats it: "I emphatically push the point that automatic, nonconscious processes pervade all aspects of mental and social life, in order to overcome what I consider dominant, even implicit, assumptions to the contrary" (p. 52).

In response to criticism that he might have overestimated the role of automatic processes in social interaction, Bargh (1997b) initially conceded that his "insinuation" that "conscious involvement is...entirely absent" from social interaction might have been "more tactical than sincere" (p. 231). Nevertheless, at the end of that same paper, he reasserted the overwhelming dominance of unconscious automaticity over conscious control: "Bloodied but unbowed, I gamely concede that the commentators did push me back from a position of 100% automaticity—but only to an Ivory© soap bar degree of purity in my beliefs about the degree of automaticity in our psychological reactions from moment to moment" (p. 246). For those who are too young to get the reference, the implication is that social cognition and behavior is 99.44% automatic.

Thus, it no surprise that Bargh has continued to assert "the unbearable automaticity of being": "[M]ost of a person's everyday life is determined not by their conscious intentions and deliberate choices but by mental processes that are put into motion by features of the environment and that operate outside of conscious awareness and guidance" (Bargh & Chartrand, 1999, p. 462).

Bargh's most recent summary of his view is simply entitled "The Automaticity of Social Life" (Bargh & Williams, 2006)—not the more modest "Automaticity *in* Social Life," which might be appropriate if automaticity were just one aspect of what goes on, but rather the sweeping implication that social life *is* automatic. Our impressions to the contrary are apparently illusions of control based on the high memorability of those occasions, roughly 0.56% of the total, when we actually have it and exercise it.

JUMPING ON THE JUGGERNAUT

Bargh is not alone in believing that automatic processes dominate experience, thought, and action, and relegating deliberate, conscious activity to the margins. Although Wegner and Schneider (1989) described a "war of the ghosts in the mind's machine" between automatic and controlled processes, they also suggested that the former tended to win out over the latter: "When we want to brush our teeth or hop on one foot, we can usually do so; when we want to control our minds, we may find that nothing works as it should. . . . Even William James, that champion of all things mental, warned that consciousness has the potential to make psychology no more than a tumbling-ground for whimsies" (p. 288).

So great was their enthusiasm for unconscious, automatic processes that these authors actually misquoted James. Here he is in full, criticizing von Hartmann (1868/1931) precisely for taking the position advocated by Wegner and Schneider—that unconscious processes rule the universe: "[T]he distinction between *the unconscious and the conscious being of the mental state* is the sovereign means for believing what one likes in psychology, and of turning what might become a science into a tumbling-ground for whimsies" (James, 1890/1980, p. 163, emphasis original). Given that this passage occurs in the context of James's 10-point critique of the notion of unconscious thought, it is clear that James considered *un*conscious processes, not conscious ones, to be the "tumbling-ground for whimsies."[1]

Nevertheless, Wegner published a book entitled *The Illusion of Conscious Will*, whose argument he summarized as follows:

> [T]he real causal mechanisms underlying behavior are never present in consciousness. Rather, the engines of causation operate without revealing themselves to us and so may be unconscious mechanisms of mind. Much of the recent research suggesting a fundamental role for automatic processes in everyday behavior (Bargh, 1997) can be understood in this light. The real causes of human action are unconscious, so it is not surprising that behavior could often arise—as in automaticity experiments—without the person's having conscious insight into its causation (D. M. Wegner, 2002, p. 97; see also D. M. Wegner & Wheatley, 1999).

Wegner's book included a diagram depicting an "actual causal path" between the "unconscious cause of thought" and "thought," and another between the "unconscious cause of action" and "action," but only an "*apparent* causal path" between thought and action.[2]

Similarly, Wilson has suggested that conscious processing may be maladaptive because it interferes with unconscious processes that are more closely tuned to the actual state of affairs in the outside world:

> . . . Freud's view of the unconscious was far too limited. When he said . . . that consciousness is the tip of the mental iceberg, he was short of the mark by quite a bit—it may be more the size of a snowball on top of that iceberg. The mind operates most efficiently by relegating a good deal of high-level, sophisticated thinking to the unconscious. . . . The adaptive unconscious

> does an excellent job of sizing up the world, warning people of danger, setting goals, and initiating action in a sophisticated and efficient manner. It is a necessary and extensive part of a highly efficient mind. (2002, pp. 6–7; for a critique, see Kihlstrom, 2004b)

The automaticity juggernaut has ranged well beyond the confines of academic psychology. Summarizing much of this research and theory, Sandra Blakeslee, a science correspondent for the *New York Times*, informed her readers that "in navigating the world and deciding what is rewarding, humans are closer to zombies than sentient beings much of the time" ("Hijacking Brain Circuits with a Nickel Slot Machine," February 19, 2002). More recently, and drawing largely on Gilbert's and Wilson's work, Malcolm Gladwell, a staff writer for the *New Yorker*, has written a trade book, *Blink*, touting the virtues of "thinking without thinking" (Gladwell, 2005).

> The part of our brain that leaps to conclusions . . . is called the adaptive unconscious, and the study of this kind of decision making is one of the most important new fields in psychology. The adaptive unconscious is not to be confused with the unconscious described by Sigmund Freud, which was a dark and murky place filled with desires and memories and fantasies that were too disturbing for us to think about consciously. This new notion of the adaptive unconscious is thought of, instead, as a kind of giant computer that quickly and quietly processes a lot of the data we need in order to keep functioning as human beings. (p. 11)

As this chapter was being finished, Gladwell's book had been on the *New York Times* nonfiction best-seller list for almost 18 months, attesting to the popularity of the concept of automaticity. It has also drawn a stern retort by Malcolm LeGault, entitled *Think: Why Critical Decisions Can't Be Made in the Blink of an Eye:*

> Predictably, as if filling a growing market niche, a new-age, feel-good pop psychology/philosophy has sprung up to bolster the view that understanding gleaned from logic and critical analysis is not all that it's cracked up to be. . . . In *Blink,* Mr. Gladwell argues that our minds possess a subconscious power to take in large amounts of information and sensory data and correctly size up a situation, solve a problem, and so on, without the heavy, imposing hand of formal thought. (p. 8)

Gladwell's book has also inspired a parody from the pseudonymous Noah Tall, entitled *Blank: The Power of Not Actually Thinking at All:*

> The part of our brain that leaps to conclusions that are reached without any thinking involved is called the *leapative concluder* or, in some circles, the

> concussive unconscious, because the unexpected hunches that suddenly slam into the brain of those who are receptive to unexpected hunches often feel exactly like being hit on the head by a heavy iron frying pan with a nonstick cooking surface.…The only reason humans have survived as long as we have despite our forgetfulness, laziness, and downright stupidity is because that tiny frying pan in our head hits us upside the unconscious when our conscious is goofing off. (Tall, 2006, pp. 7–8)

THE THIRD-AND-A-HALF DISCONTINUITY?

Experimental evidence indicates that automatic processes play some role, under some conditions, in social cognition and behavior. On this much we can agree. But what might be called the doctrine of automaticity goes way beyond such restricted conclusions to assert that automatic processes pervade human experience, thought, and action; conscious awareness is largely an afterthought; and conscious control is an illusion. Humans are, in this view, a special class of zombies, virtual automatons who are conscious, as La Mettrie had argued, but for whom consciousness plays little or no functional role in thought and action. The purpose of consciousness is to erect personal theories about why things happen as they do, and why we do what we do. But, on this view, consciousness is largely irrelevant to what actually goes on. Bargh puts the point concisely: "As Skinner argued so pointedly, the more we know about the situational causes of psychological phenomena, the less need we have for postulating internal conscious mediating processes to explain these phenomena" (1997a, p. 1).

Of course, the progress of science will by its very nature correct popular misunderstandings of how the world works, and occasionally reveal surprising, even unpleasant, truths about ourselves. Sigmund Freud famously situated himself in line with Copernicus, who taught us that Earth is not at the center of the universe, and Darwin, who taught us that humans are creatures of nature just like any other. For Freud, the third blow against "human megalomania" was his discovery (as he claimed it was) that conscious experience, thought, and action was determined by unconscious, primitive drives: "[H]uman megalomania will have suffered its third and most wounding blow from the psychological research of the present time which seeks to prove to the ego that it is not even master in its own house, but must content itself with scanty information of what is going on unconsciously in the mind" (Freud, 1915–1917/1961–1963, p. 285; see also Bruner, 1958).

Bargh has explicitly situated himself in this line of scientific progress, substituting for Freud's irrational "monsters from the Id" a view of humans as operating not necessarily irrationally, but whether rational or not, operating mostly on automatic pilot, uninfluenced by conscious deliberation: "[W]e are not as conscious, or as free, as we thought we were" (Bargh, 1997a, p. 52). Henceforth, we must live with "the unbearable automaticity of being" (Bargh & Chartrand, 1999).

Like Bargh, Wegner and Smart (1997) also replaced Freud's third discontinuity, substituting automaticity for irrationality. For the record, there also seems to be a fourth discontinuity, between humans and machines, which some visionaries, like Mazlish (1993) and Kurzweil (1999) see as being erased by advances in artificial intelligence. Of course, the idea that humans are simply machines—if machines made of meat—is entirely consonant with the idea that human experience, thought, and action are the product of unconscious processes operating automatically.

It would be one thing if the doctrine of automaticity were backed by sound scientific evidence. Then, we would have no choice but to shrug our shoulders, cast off our sentimental beliefs in conscious control, and free will, and find some way to bear "the unbearable automaticity of being," just as we have learned to live with the knowledge that Earth is not the center of the universe and that humans are not the products of Special Creation. But in fact, the doctrine of automaticity is not true—or, at least, it is not backed by sound scientific evidence. There are at least three reasons for thinking that the third discontinuity, at least the one erased by Bargh and Wegner (never mind Freud) is not quite ready to be expunged.

The first reason, paradoxically, is that the theoretical underpinnings of the concept of automaticity have begun to unravel (G. D. Logan, 1997; Moors & DeHouwer, 2006; Pashler, 1998). In particular, the resource theories of attention on which the concept was originally based have come into question. For example, there does not seem to be a single pool of attentional resources. Nor does even extensive practice with a task render its performance effortless. There is even some data that suggests that attentional capacity is not limited—at least, that its limits are very wide indeed. As noted earlier, alternative theories of automaticity have been proposed, particularly based on memory rather than attention memory. These revisionist theories preserve the legitimacy of the concept of automaticity but tend to undercut the various features by which automatic processes are recognized. So, for example, in J. R. Anderson's (1992) proceduralization view, automatic processes are engaged only when an appropriate cue is presented in the context of a particular goal state; and in Logan's (2002) instance-based theory, automatic processes are evoked only if the subject has the appropriate mental set. Further, once evoked, automatic processes do not necessarily proceed to their conclusion unimpeded,in a ballistic fashion.

One response to this state of affairs is to abandon the assumption that the distinction between automatic and controlled processes is a qualitative, all-or-none matter; rather, it is argued, automaticity varies by degrees (Bargh, 1989, 1994). This response is fine, and almost certainly correct, but it has the unfortunate consequence of making it difficult to know precisely when a process is automatic, and when it is not. What happens, for example, if a process seems to run off unintentionally, but nevertheless consumes attentional capacity? And, of course, the concession that some tasks are performed *more or less* automatically undercuts the fundamental message of "the automaticity of social life" (Bargh & Williams, 2006).

Moreover, it should be noted that the shift to a continuous view of automaticity has been accompanied by a certain slippage in the operationalization of the concept in psychological experiments. For example, in his earliest research Bargh employed a dichotic listening task (Bargh, 1982) or parafoveal presentation (Bargh & Pietromonaco, 1982) in an effort to conform to a relatively strict operational definition of automaticity. Similarly, Fazio et al. (1986) and Devine (1989) employed extremely short prime-target intervals, in an attempt to prevent their subjects from employing controlled processes. But in more recent work, such strictures are often abandoned. For example, Bargh and his colleagues have presented words in subjects' clear view, and asked them to pronounce them (Bargh, Chaiken, Raymond, & Hymes, 1996), or to assemble them into sentences (Bargh et al., 1996)—tasks that would seem to involve conscious processing. Granted, in these cases the subjects were not specifically instructed to process the relevance of the words to certain attitudes and stereotypes, thus approximating the unintentional nature of automatic processing. But this reliance on only a single feature is a considerable departure from the concept of automaticity as it was originally set out in cognitive psychology.

In fact, within social psychology the concept of automaticity seems to be invoked whenever subjects engage in processing that is incidental to the manifest task set for them by the experimenter—whether this is shadowing text, detecting visual stimuli, pronouncing words, or assembling sentences. But just because something is done incidentally does not necessarily mean that it has been performed unintentionally, much less automatically. In many situations, subjects may have plenty of processing capacity left over, after the manifest task has been performed, and they may use some of it, quite deliberately, to perform other tasks that interest them—such as critically analyzing the experiment's cover story, or speculating about the experimenter's true purposes (Orne, 1962, 1973).

Most critically, the social-psychological literature on automaticity rarely contains any actual comparison of the strength of automatic and controlled processes. These were features of some of the earliest experiments on automaticity: In studies already described, for example, Fazio et al. (1986), and Devine (1989) also employed relatively long prime-target intervals in their experiments, in an attempt to compare the effects automatic and controlled processing. Within cognitive psychology, there has been considerable interest in developing techniques such as the process-dissociation procedure (PDP; L. L. Jacoby, 1991) to directly compare the contributions of automatic and controlled processes to task performance. For example, Jacoby and his colleagues (1997) showed convincingly that successful recognition was mediated mostly by controlled retrieval in young subjects, but mostly by automatic familiarity in the elderly. The PDP has its critics (e.g., Curran & Hintzman, 1995), but the point is that cognitive psychologists tend to assume that both automatic and controlled processes contribute to task performance, and try to disentangle them. By contrast, an increasingly popular view within social psychology is that automatic processes dominate, and controlled processes are largely irrelevant.

In fact, attempts to use the PDP or some similar procedure to directly compare the strengths of automatic and controlled processes in some social-psychological task are vanishingly rare. As of this writing, fewer than a dozen such papers had appeared in personality and social psychology journals, out of the total corpus of automaticity papers described above. In one recent attempt, Uleman and his colleagues (Uleman, Blader, & Todorov, 2005, Experiment 1) presented subjects with photographs of target individuals, paired with descriptions of their behaviors; immediately afterward, or after 20-minute or 2-day delays, they were asked to rate the targets on a variety of personality traits. In an "inclusion" condition, the subjects were told that the behavior descriptions were informative about the targets' traits, and they should consider them; in an "exclusion" condition, they were told that the behavior descriptions were irrelevant, and they should ignore them. This "method of opposition" pits automatic and controlled processes against each other, and permits an estimate of the contributions of both kinds of processes to performance on the trait-rating task. In fact, the experiment showed that controlled processes dominated automatic ones on the immediate trait ratings; and there was almost a perfect balance between them on the ratings made after both short and long delays. Although it is true, as Uleman et al. note, that delay reduced the impact of controlled processing on task performance, results like this are far from showing that automatic processes are *more* important than controlled processes—much less that controlled, conscious processes are an afterthought, irrelevant to human experience, thought, and action.

Similarly, Payne et al. (2005) employed a variant of the PDP to test two models of the relation between controlled and automatic processing in the "weapon identification effect," by which an ambiguous object is more likely to be identified as a weapon if it is held by a Black than by a White person. Payne (2001) had earlier shown that both automatic and controlled processes play a role in this effect, and that the influence of automatic processes was magnified under conditions of a response deadline. But in a reanalysis of this data, Payne et al. showed that the effect was dominated by controlled processes, and automatic processes played a relatively subordinate role, biasing judgments only when the controlled process was absent. In fact, their analysis specifically rejected the widely popular two-stage view of automaticity in stereotyping, in which stereotypes are automatically activated and must be overcome by controlled processes (see also Payne & Stewart, 2007). At any rate, it should come as no surprise that automatic processes play an especially important role when subjects are required to respond within 500 msec of a stimulus.

THE ALLURE OF AUTOMATICITY

It is one thing to assert that automatic processes play a role in social interactions, along with controlled processes, with the proviso that some automatic processes are more automatic than others. It is another thing entirely to embrace and promote the idea that automatic processes dominate human experience, thought, and action to the virtual exclusion of everything else. Although there is plenty of evidence that automatic processes play some role in social cognition and behavior, as they probably do in almost every aspect of human performance, nothing in any experimental demonstration of automaticity demands such a sweeping inference.

So why are some social psychologists inclined to take this further, empirically unjustified, and logically unnecessary, step? Perhaps, if the step is not motivated by empirical data, then it is motivated by something closer to the *a priori* and *quasi-metaphysical* reasons criticized by James more than a century ago.

Partly, the enthusiasm for automaticity seems to reflect a reaction against the "cognitive revolution" in social psychology, with its (tacit) view of social interaction as mediated by conscious, deliberate, rational thought—as reflected, for example, in balance theory (Heider, 1946, 1958), cognitive consistency theory (Festinger, 1957; see also Abelson et al., 1968), cognitive algebra (N. H. Anderson, 1974), and early formulations of attribution theory (Kelley, 1967). It is also probably not an accident that social psychologists' interest in automaticity began to develop at roughly the same time as the "affective counterrevolution" emerged in social psychology, with its view of affective states as automatically generated by environmental stimuli, independent of cognitive analysis (Zajonc, 1980, 1984). In fact, Zajonc (1999) has explicitly connected the two themes of automaticity and emotion.

Then, too, the biologization of social psychology may contribute to a reduced role for conscious control in theories of social interaction. To the extent that the reasons for particular patterns of social interaction are to be found in "selfish" genes whose only goal is their own reproduction (Dawkins, 1976), there seems to be little room for the kind of conscious, deliberate thought that we commonly associate with human intelligence. So, too, if social interaction is driven by mental and behavioral instincts that we share with our nonhuman ancestors (Barkow, Cosmides, & Tooby, 1992; Buss, 1999). Finally, social neuroscience (Cacioppo, Berntson, & McClintock, 2000) can, unless we are careful, veer into a reductionism that leaves conscious thought and other aspects of commonsense "folk psychology" entirely out of the explanation of behavior (Churchland, 1986).

Although each trend entails risks, both the emergence of an affective psychology paralleling cognitive psychology and an interest in the neural and other biological underpinnings of social interaction should be seen as positive developments within social psychology. But there also seems to be a darker side to the current interest in automaticity. Currently, mainstream social psychology is characterized by a focus on judgment error, normative violations, and other aspects of social *mis*behavior (Krueger & Funder, 2004). Although it may be true (or at least arguable) that science learns more from counterintuitive findings that undercut commonsense "folk psychology," it is also true that this emphasis on the negative can degenerate into what might be called a "People Are Stupid" school of psychology (Kihlstrom, 2004a). That is, as we go about the ordinary course of everyday living, we do not think very hard about anything, and rely on biases, heuristics, and other processes that lead us into judgmental error (e.g., Nisbett & Ross, 1980; Ross, 1977; see also Gilovich, 1991). In this view, the evidence for irrationality consists not just in demonstrations of various heuristics and biases in judgment, because some of these might merely be evidence of bounded rationality (Simon, 1957), but also evidence of unconscious, automatic processes. It is not just that we do not think too hard about things; we also do not pay too much attention to what is going on around us or to what we are doing (Gilbert & Gill, 2000).

Nor do we know too much about why we do what we do (Nisbett & Wilson, 1977; T. D. Wilson & Stone, 1985; W. R. Wilson, 1979). Thought and behavior just happens, automatically, in response to environmental stimuli, and our belief that we control what we think and do amounts to little more than an illusion, an after-the-fact rationale. In fact, our attempts to consciously control our experience, thought, and action typically backfire (D. M. Wegner, 1989), and we would be better off if we relied on automatic processes (T. D. Wilson, 2002).

Also on the dark side is a long-standing, but again largely unspoken, alliance between social psychology and behaviorism (Zimbardo, 1999). Just as Watson (1913, 1919) and Skinner (1938, 1953, 1977, 1990) viewed behavior as under the control of environmental stimuli, so social psychology has historically been defined as concerned with the influence of the social situation on the individual's experience, thought, and action. Floyd Allport (1924), in his pioneering text on social psychology, adopted an expressly behavioristic stance, interpreting social behavior either as the response to the stimulus of another person's behavior or as a stimulus to another person's response. The behaviorist emphasis on the situation was codified by Gordon Allport 30 years later (1954), when he defined social psychology as the study of "how the thought, feeling, and behavior of individuals are influenced by the actual, imagined, or implied presence of other human beings" (p. 1).[3]

We can see the behaviorist emphasis on social behavior as response to environmental stimuli in the "Four As" of social psychology—aggression, altruism, attitude change, and attraction; in the classic studies of social facilitation and other aspects of social impact, conformity, persuasion; and elsewhere on almost any randomly selected page of a typical social psychology textbook. The doctrine of situationism is so firmly entrenched in social psychology that Ross and Nisbett (1991) identified "the principle of situationism" as the first leg of "the tripod on which social psychology rests" (p. 8). Although the cognitive perspective in social psychology that emerged in the 1960s often stressed the importance of the *perceived* situation, in fact many of the classic studies in the field made little or no reference to the internal cognitive processes by which individuals constructed the mental representations of the situation that actually governed their behavior.

As Berkowitz and Devine (1995) have noted, all of this classic literature can be reinterpreted in terms of the automatic elicitation of feelings, thoughts, and actions by environmental stimuli. Wegner and Bargh (D. M. Wegner & Bargh, 1998) agree:

> Classic social psychology…makes people appear to be automatons. The situational influences on behavior investigated in these [classic] studies were (a) *unintended* on the part of the individual, (b) not something of which the person was *aware*, (c) a response to the situation occurring before the individual had a chance to reflect on what to do (i.e., *efficient*) or (d) *difficult to control* or inhibit even when the person is cognizant of the influence. As it happens, these are characteristics of automatic psychological processes, not of conscious control, and comprise a handy working definition of automaticity. (p. 447)

Of course, it should be noted that these classic experiments were all conducted before the concept of automaticity emerged in cognitive psychology. Therefore, we do not really know whether the effects they yielded were unintended, unaware, efficient, or difficult to inhibit.

A recent overview of social psychology intended for neuroscientists made the connection between situationism and automaticity even clearer:

> If a social psychologist was going to be marooned on a deserted island and could only take one principle of social psychology with him it would undoubtedly be "the power of the situation." All of the most classic studies in the early days of social psychology demonstrated that situations can exert a powerful force over the actions of individuals.…
>
> If the power of the situation is the first principle of social psychology, a second principle is that people are largely unaware of the influence of situations on behavior, whether it is their own or someone else's behavior. (Lieberman, 2005, p. 746)

The reason that people are blind to situational influences is that situational influences operate automatically and unconsciously.

Bargh himself has clearly connected behaviorism, situationism, and automaticity with the problem of free will (1997a, p. 1):

> Now, as the purview of social psychology is precisely to discover those situational causes of thinking, feeling, and acting in the real or implied presence of other people . . . , it is hard to escape the forecast that as knowledge progresses regarding psychological phenomena, there will be less of a role played by free will or conscious choice in accounting for them. In other words, because of social psychology's natural focus on the situational determinants of thinking, feeling, and doing, it is inevitable that social psychological phenomena will be found to be automatic in nature.

The automaticity juggernaut is not strictly a return to stimulus-response behaviorism, because it agrees that cognitive processes mediate between stimulus and response. Thus it is able to maintain a superficial allegiance to cognitivism while harkening back to a radical situationism. If the cognitive processes underlying interpersonal relations behavior are automatically triggered by environmental cues, then behavior is determined by the environment; if social behavior is not absolutely automatic, at least not *too much* thought has gone into it. Inspired by the late Susan Sontag, we can think of this as behaviorism with a cognitive face.

ARE WE AUTOMATONS AFTER ALL?

Although the cognitive revolution made the study of consciousness respectable again (Hilgard, 1980), the topic of consciousness has always made some psychologists (and other cognitive scientists) nervous, resulting in what Flanagan (1992) has dubbed *conscious shyness.* In part, conscious shyness reflects a kind of positivist reserve, itself a holdover from behaviorism, which prefers behavior over self-reports as the data for psychology; in part, it reflects a strategic preference for approaching consciousness obliquely, through studies of perception, memory, and the like that do not expressly evoke the concept of consciousness. But there is more too it than that. In Flanagan's view, *conscious inessentialism,* or the idea that conscious awareness and control is not necessary for many aspects of cognition, feeds the *epiphenomenalist suspicion* that consciousness plays no causal role in behavior after all. In this view, we may be conscious zombies, but we are zombies nonetheless.

By embracing the concept of automaticity, we can admit that we have consciousness, and even search for its neural correlates, without also admitting that consciousness has anything to do with causing our behavior. As noted earlier,

D. M. Wegner (2002) has vigorously argued that conscious control is an illusion, and that our conscious intentions are previews of action, not the causes of it. As he puts it, "This is the way it needs to be for progress in the explanation of human psychology. The agent self cannot be a real entity that causes actions, but only a virtual entity, an apparent mental causer" (2005, p. 23). This quote makes it clear that the automaticity juggernaut is fueled by pre-theoretical ideological commitments, rather than any empirical findings—not just to the doctrine of situationism, or to the behaviorist viewpoint, but to a particular view of what science is, and what kinds of explanations a scientific theory allows.

Indeed, epiphenomenalism, in turn, links to a perennial problem for psychology, and indeed for all the social sciences, which is the question of free will and determinism (Rogers & Skinner, 1956). To some theorists, the idea that consciousness actually plays a causal role in behavior seems to violate the fundamental assumption of the scientific enterprise—that every event has a physical cause, and that human—or, for that matter, superhuman—agency has no place in scientific explanation. Given the choice between adhering to the assumption of determinism and taking consciousness seriously, some scientists choose the former, construing thought and action as automatic and consciousness as epiphenomenal, without causal efficacy. Thus, Bargh and Ferguson (2000) write that automaticity succeeded where behaviorism failed, solving the problem of free will by showing how behavior could be determined by the stimulus environment after all:

> [T]he same higher mental processes that have traditionally served as quintessential examples of choice and free will—such as goal pursuit, judgment, and interpersonal behavior—have been shown recently to occur in the absence of conscious choice or guidance. It would seem, therefore, that the mid-century failure of behaviorism to demonstrate the determinism of complex higher order human behavior and mental processes occurred not because those processes were not determined but rather because behaviorists denied the existence of the necessary intraindividual, psychological explanatory mechanisms . . . mediating between the environment and those higher processes. . . .
>
> [T]he failure of behaviorism in no way constituted the failure of determinism. We . . . present the case for the determinism of higher mental processes by reviewing the evidence showing that these processes, as well as complex forms of social behavior over time, can occur automatically, triggered by environmental events and without an intervening act of conscious will or subsequent conscious guidance. (p. 926)

One is tempted to ask whether we really had a cognitive revolution in psychology for this—to learn that Skinner had it right after all, that we really are all under the control of environmental events, and that all he missed was the wiring diagram that connects stimulus with response.

For Wegner (2002), as for Bargh and Ferguson (2000), it seems that automaticity is the key to the scientific status of psychology itself. Automaticity does more than demystify unconscious mental life: It permits us to bypass the will (Bargh, 2005), and allow psychology to adopt the pinball determinism of classical physics. Bargh, Wegner, and others, faced with an apparent conflict between free will and determinism, choose determinism, and automaticity is a means for doing just that. At the same time, this may be a false choice. Certainly there is nothing in the scientific evidence concerning the role of automatic processes in social behavior that would compel us to choose automaticity over control.

As Searle (1992, 2000a, 2000b, 2001a, 2001b) has argued, whenever we are confronted by a choice between two equally compelling beliefs, such as our experience of free will and our scientific commitment to determinism, it is likely that the choice has been poorly framed to begin with. Perhaps we need to jettison the notion of free will as a sentimental component of folk psychology that must be abandoned in the face of the progress of science. Or perhaps the proper stance is to accept the experience of conscious will as valid, and try to explain how free will can enter into the causal scheme of things in a material world of neurons, synapses, and neurotransmitters. The choice is ours to make: Our choice will determine whether we will have a science of the mind worth having.

NOTES

1. James is often misunderstood here. James identified consciousness with thinking, defined broadly to include feeling as well. Accordingly, “unconscious thought” struck him as an oxymoron. Still, his studies of hypnosis and hysteria led him to thinking and feeling could occur outside of phenomenal awareness. Hence his metaphor of the stream of consciousness, which could be divided, with one stream out of contact with the other—an idea subsequently revived in Hilgard’s (1977) neodissociation theory of divided consciousness.

2. For critical analyses of Wegner’s arguments, see Kihlstrom (2004c), and other commentaries accompanying the book précis, as well as Wegner’s reply, published in *Behavioral & Brain Sciences* (2004), Vol. 27, No. 6.

3. This is ironical in the extreme, given that Gordon Allport and Skinner were such vigorous adversaries that the Harvard psychology department had to be split in order to accommodate them (E. R. Hilgard, 1987).

REFERENCES

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (Eds.). (1968). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally.

Allport, F. H. (1924). *Social psychology.* Boston: Houghton Mifflin.
Allport, G. W. (1954). The historical background of social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1, pp. 1–46). New York: Random House.
Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105*(2), 165–180.
Anderson, N. H. (1974). Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1–101). New York: Academic Press.
Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology, 43*, 425–436.
Bargh, J. A. (1984). Automatic and conscious processing of social information. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–43). Hillsdale, NJ: Erlbaum.
Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: Causes and consequences for judgment, emotion, and behavior* (pp. 3–51). New York: Guilford.
Bargh, J. A. (1990). Auto-motives: Preconscious determinants of social interaction. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition* (pp. 93–130). New York: Guilford.
Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ: Erlbaum.
Bargh, J. A. (1997a). The automaticity of everyday life. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.
Bargh, J. A. (1997b). Reply to the commentaries. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 231–246). Mahwah, NJ: Erlbaum.
Bargh, J. A. (2005). Bypassing the will: Toward demystifying the nonconscious control of social behavior. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 37–58). New York: Oxford University Press.
Bargh, J. A. (Ed.) (2007). *Social psychology and the unconscious: The automaticity of higher mental processes.* New York: Psychology Press.
Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32*, 104–128.
Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462–479.
Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality & Social Psychology, 71*, 230–244.
Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126*(6 Special Issue), 925–945.
Bargh, J. A., & Gollwitzer, P. M. (1994). Environmental control of goal-directed action: Automatic and strategic contingencies between situations and behavior.

In *Nebraska Symposium on Motivation* (Vol. 41, pp. 71–124). Lincoln: University of Nebraska Press.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality & Social Psychology, 43*, 437–449.

Bargh, J. A., & Uleman, J. S. (1989). Introduction. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: Causes and consequences for judgment, emotion, and behavior* (pp. xiii–xxvi). New York: Guilford.

Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science, 15*(1), 1–4(4).

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Berkowitz, L., & Devine, P. G. (1995). Has social psychology always been cognitive? What is "cognitive" anyhow? *Personality & Social Psychology Bulletin, 21*(7), 696–703.

Bruner, J. S. (1958). The Freudian concept of man and the continuity of nature. *Daedalus, 87*, 77–44.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind.* Boston: Allyn and Bacon.

Cacioppo, J. T., Berntson, G. G., & McClintock, M. K. (2000). Multilevel integrative analyses of human behavior: Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin, 126*(6), 829.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology.* New York: Guilford.

Churchland, P. S. (1986). *Neurophilosophy : Toward a unified science of the mind-brain.* Cambridge, MA: MIT Press.

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 531–547.

Dawkins, M. S. (1976). *The selfish gene.* Oxford: Oxford University Press.

Devine, P. G. (1989). Automatic and controlled processes in prejudice: The role of stereotypes and personal beliefs. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 181–212). Hillsdale, NJ: Erlbaum.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality & Social Psychology, 50*, 229–238.

Ferguson, M. J., & Bargh, J. A. (2004). How social perception can automatically influence behavior. *Trends in Cognitive Sciences, 8*(1), 33–39(37).

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Flanagan, O. (1992). *Consciousness reconsidered.* Cambridge, MA: MIT Press.

Freud, S. (1961–1963). Introductory lectures on psychoanalysis. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 16–17). London: Hogarth Press. (Original work published 1915–1917)

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford.

Gilbert, D. T., & Gill, M. J. (2000). The momentary realist. *Psychological Science, 11*, 394–398.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York: Free Press.

Gladwell, M. (2005). *Blink: The power of thinking without thinking.* Boston: Little, Brown.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality & Social Psychology, 74*, 1464–1480.

Hartmann, E. V. (1931). *Philosophy of the unconscious: Speculative results according to the inductive method of physical science.* London: Routledge and Kegan Paul. (Original work published 1868)

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.). (2005). *The new unconscious.* New York: Oxford University Press.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21*, 107–112.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Higgins, E. T., & King, G. (1981). Accessibility of social constructs: Information-processing consequences of individual and contextual variability. In N. Cantor & J. F. Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 69–121). Hillsdale, NJ: Erlbaum.

Hilgard, E. R. (1977). *Divided consciousness: Multiple controls in human thought and action.* New York: Wiley-Interscience.

Hilgard, E. R. (1980). Consciousness in contemporary psychology. *Annual Review of Psychology, 31*, 1–26.

Hilgard, E. R. (1987). *Psychology in America: A historical survey.* New York: Harcourt Brace Jovanovich.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language, 13*, 513–541.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In *Scientific approaches to consciousness* (pp. 13–47). Hillsdale, NJ: Erlbaum.

James, W. (1890/1980). *Principles of psychology.* Cambridge, MA: Harvard University Press.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. N. Levine (Ed.), *Nebraska Symposium on Motivation 1967.* Lincoln, NB: University of Nebraska Press.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237*(4821), 1445–1452.

Kihlstrom, J. F. (2004a). Is there a "People Are Stupid" school in social psychology? [Commentary on "Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition" by J. I. Krueger and D. C. Funder]. *Behavioral & Brain Sciences, 27*, 348–349.

Kihlstrom, J. F. (2004b). Is your unconscious smarter than you are? [Book review of *Strangers to ourselves: Discovering the adaptive unconscious* by T. D. Wilson]. *PsycCRITIQUES, 49*(Supplement 14), Article 99.

Kihlstrom, J. F. (2004c). An unwarrantable impertinence [Commentary on *The Illusion of Conscious Will* by D. M. Wegner]. *Behavioral & Brain Sciences, 27*, 666–667.

Krueger, J. I., & Funder, D. C. (2004). Toward a balanced social psychology: Causes, consequences and cures for the problem-seeing approach to social behavior and cognition. *Behavioral & Brain Sciences*, 27, 313–376.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence.* New York: Viking.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

La Mettrie, J. O. (1748/1749). *Man a machine.* London: W. Owen.

Langer, E., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology, 36*, 635–642.

Langer, E. J. (1978). Rethinking the role of thought in social interaction. In J. H. Harvey, W. J. Ickes & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2, pp. 35–58). Potomac, MD: Erlbaum.

LeGault, M.R. (2006). *Think: Why crucial decisions can't be made in the blink of an eye.* New York: Threshold.

Lieberman, M. D. (2005). Principles, processes, and puzzles of social cognition: An introduction for the special issue on social cognitive neuroscience. *NeuroImage, 28*, 745–756.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Logan, G. D. (1997). The automaticity of academic life: Unconscious applications of an implicit theory. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 157–179). Mahwah, NJ: Erlbaum.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109*(2), 376–400.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203.

Mazlish, B. (1993). *The fourth discontinuity: The co-evolution of humans and machines.* New Haven, CT: Yale University Press.

Moors, A., & DeHouwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297–326.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., & Wilson, D. S. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–253.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality & Social Psychology Bulletin, 31*, 166–180.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.

Orne, M. T. (1973). Communication by the total experimental situation: Why it is important, how it is evaluated, and its significance for the ecological validity of findings. In P. Pliner, L. Krames, & T. Alloway (Eds.), *Communication and affect* (pp. 157–191). New York: Academic.

Park, D. C. (1999). Acts of will? *The American Psychologist, 54*(7), 461.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: The MIT Press.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality & Social Psychology, 81*, 181–192.

Payne, B. K., Jacoby, L. L., & Lambert, A. J. (2005). Attitudes as accessibility bias: Dissociating automatic and controlled processes. In R. R. Hassin, J. S. Uleman & J. A. Bargh (Eds.), *The new unconscious* (pp. 393–420). New York: Oxford University Press.

Payne, B. K., & Stewart, B. D. (2007). Automatic and controlled components of social cognition: A process dissociation approach. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 293–315). New York: Psychology Press.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). New York: Wiley.

Rogers, C. R., & Skinner, B. F. (1956). Some issues concerning the control of human behavior. *Science, 124*, 1057–1066.

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). San Diego: Academic Press.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1–66.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

Searle, J. R. (2000a). Consciousness. *Annual Review of Neuroscience,* 23, 557–578.

Searle, J. R. (2000b). Consciousness, free action and the brain. *Journal of Consciousness Studies, 7(10), 3–22.*

Searle, J. R. (2001a). Free will as a problem in neurobiology. *Philosophy, 76,* 491–514.

Searle, J. R. (2001b). *Rationality in action.* Cambridge, MA: MIT Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127–190.

Shiffrin, R. M., & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review, 91*(2), 269–276.

Simon, H. A. (1957). *Models of man: Social and rational.* New York: Wiley.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* New York: Appleton-Century.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism, 7,* 1–10.

Skinner, B. F. (1990). Can psychology be a science of mind? *American Psychologist, 45,* 1206–1210.

Smith, E. R., & Miller, F. D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review, 85,* 355–362.

Tall, N. (2006). *Blank: The power of not actually thinking at all.* New York: Harper.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 249–288). New York: Academic Press.

Uleman, J. S., & Bargh, J. A. (Eds.). (1989). *Unintended thought.* New York: Guilford Press.

Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 362–392). New York: Oxford University Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158–177.

Watson, J. B. (1919). *Psychology from the standpoint of a behaviorist.* Philadelphia: Lippincott.

Wegner, D. M. (1989). *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control.* New York: Penguin.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wegner, D. M. (2005). Who is the controller of controlled processes? In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 19–36). New York: Oxford University Press.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 1, pp. 446–496). Boston: McGraw-Hill.

Wegner, D. M., & Schneider, D. J. (1989). Mental control: The war of the ghosts in the machine. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: The limits of consciousness, intention, and control* (pp. 287–305). New York: Guilford.

Wegner, D. M., & Smart, L. (1997). Deep cognitive activation: A new approach to the unconscious. *Journal of Consulting & Clinical Psychology, 65,* 984–995.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation—Sources of the experience of will. *American Psychologist, 54*(7), 480–492.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Belknap Press of Harvard University Press.

Wilson, T. D., & Stone, J. I. (1985). Limitations of self-knowledge: More on telling more than we can know. In P. Shaver (Ed.), *Review of personality and social psychology* (Vol. 6, pp. 167–183). Beverly Hills, CA: Sage.

Wilson, W. R. (1979). Feeling more than we can know: Exposure effects without learning. *Journal of Personality & Social Psychology, 37*(6), 811–821.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist, 39*, 117–123.

Zajonc, R. B. (1999). One hundred years of rationality assumptions in social psychology. In A. Rodrigues & R. V. Levine (Eds.), *Reflections on 100 years of social psychology* (pp. 200–214). New York: Basic Books.

Zimbardo, P. G. (1999). Experimental social psychology: Behaviorism with minds and matters. In A. Rodrigues & R. V. Levine (Eds.), *Reflections on 100 years of experimental social psychology* (pp. 135–157). New York: Basic Books.

9 The Hazards of Claiming to Have Solved the Hard Problem of Free Will

Azim F. Shariff
Jonathan Schooler
Kathleen D. Vohs

She raised one hand and flexed its fingers and wondered how this thing, this fleshy spider on the end of her arm, came to be hers, entirely at her command. She bent her finger and straightened it. The mystery was in the moment before it moved, the dividing instant between not moving and moving, when her intention took effect. It was like a wave breaking. . . . There was no stitching, no seam, and yet she knew that behind the smooth continuous fabric was the real self—was it her soul?—which took the decision to begin movement and gave the final command.

—Ian McEwan, *Atonement*

"You," your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. Who you are is nothing but a pack of neurons.

—Francis Crick, *The Astonishing Hypothesis*

Nothing approaching the truth has yet been said on this subject.

—Thomas Nagel, *The View From Nowhere*

Up in the ivory heights, the free will debate has been raging for centuries—first in philosophy, and more recently, and perhaps more fiercely, in the brain sciences. It has been a divisive argument between the allegedly naïve position that people are in conscious control of their actions and the counterintuitive position that this experience of free will is illusory and people are automatons being pushed around by the compendium of known forces in a physical world.

For all the ostensible importance of such a question, the debate has had little to no impact outside of academia. For most people, the apparent volition behind their own behavior is satisfactory. What people want to happen, happens, and people generally assume that their conscious responses caused the outcome (Wegner, 2005). The rejoinder espoused by many scientists is that the connection between volition and action is merely correlational, and not causational. But this is held by many to be a purely academic argument that is at best beside the point and at worst absurd. Meanwhile, many philosophers of free will are satisfied with the compatibilist understanding that any type of free will worth wanting, such as freedom from coercion, compulsion, and political oppression, is unthreatened by scientific findings (Dennett, 2003). Although from time to time, people have wondered about whether this academic debate might have repercussions for the lay public (Breer, 1989; Skinner, 1971), such musings have been largely speculative, as no empirical evidence one way or the other has been brought to bear on the issue.

A recent set of studies, reviewed below, provides preliminary evidence that exposure to academic claims regarding the absence of free will can have an impact on moral action (Vohs & Schooler, 2008). Having one's traditional understanding of free will disturbed by the determinist argument seems to encourage a form of moral laxity. Contrary to the view that discussions of free will are largely academic, this work suggests that the belief in free will, be it justified or mistaken, affects behavior. Although it may be true that free will is an impotent epiphenomenon, the belief in free will can have real and potent consequences. Should the illusory free will position advanced in academic circles enjoy popular support among the lay public, it may be accompanied by larger social implications. The message that there is no free will may go from being understood as "nothing is controllable" to "everything is permitted." Again, regardless of the actual status of free will, a scientifically backed repudiation of it may encourage debauched behavior.

Put simply, the question of free will matters. And it matters not only to scientists in labs and philosophers in armchairs, but to the way that people live their lives. In a free society, neither scientists nor philosophers should be asked to suppress their views for fear of the possible social ramifications that the expression of such views might have. Nevertheless, if science is to be used as foundation upon which to promote claims that may have social impact, then the soundness of those claims deserves particular scrutiny. In this chapter, we consider the question of whether science has reached the point at which the notion of free will must be dismissed, as some authors have argued (e.g., Crick, 1994; Pinker, 2002). Drawing on parallels between the challenges of conceptualizing free will and challenges in resolving the "hard problem" versus easy problems of consciousness (Chalmers, 1995), we argue that there exist two general classes of problems associated with the question of free will. The easy problems are those that have recently been posed within psychology—issues of automaticity, timing and backward referral, the neural systems involved, and so forth. These problems (reviewed in depth below) are by no means trivial, but they are akin to the easy problems of consciousness, which is to say that they are problems that can conceivably be solved using methods that are currently at the disposal of the scientific community. The hard problem of free will is different. Intricately related to the hard problem of consciousness, the hard problem of free will represents *the* core problem of conscious free will: Does conscious volition impact the material world? In other words, can phenomenal experiences translate into a physical events? And if so, how?

Although making this division should face similar criticisms as the ones leveled against Chalmers (e.g., Dennett, 1995) 10 years ago, there is particular need to do so. Without a clear distinction between the classes of problems, it is easy to confuse progress toward the easy problem as evidence pertaining to the hard problem. By clarifying which class of problems each scientist is investigating, confusion will be muted, and scientists can maintain as focus on the scope of their conclusions.

THE HARD PROBLEM OF CONSCIOUSNESS

Before distinguishing between the hard and easy problems of free will, it will be useful to first review this distinction as Chalmers (1995) applied it to consciousness. According to Chalmers, the easy problems of consciousness involve those that can be tackled with the standard methods of cognitive science and can be explained with computational and neural mechanisms. These include, but are not limited to, questions such as how attention modulates conscious experience, the difference between wakefulness and sleep, and the reportability of mental states. Although these problems are challenging, they are labeled as "easy" because they can be conceptualized within neurocognitive explanatory systems, and tested accordingly.

Chalmers separates these easy problems from the hard problem of consciousness—how the brain gives rise to the subjective experience that humans recognize as consciousness. This problem is hard because it cannot be conceptualized within standard neurocognitive explanatory systems. Current neurocognitive systems offer no principled way to distinguish between a system that has experience from one that does not. The elusiveness of subjectivity is illustrated by the hypothetical notion of zombies: individuals who are organized and behave just like us but are devoid of subjective experience. Because subjective experience is (obviously) entirely subjective, there is no principled way to distinguish a zombie that claims to have subjective experience but does not, from a person who genuinely experiences consciousness. A similar problem occurs in determining whether nonhuman species experience consciousness. How can humans know what it is like to be a bat (Nagel, 1974), or, for that matter, a cockroach, or an amoeba, or a computer? Where do humans draw the line, and on what grounds? The hard problem of consciousness is hard because, in contrast to the easy problems, it is unclear what type of evidence would constitute a solution for it. Indeed, some have suggested that it is beyond the capacity of the human brain to even fathom a solution (McGinn, 1999).

Countless articles have been written about the hard problem of consciousness as articulated by Chalmers and his predecessors (e.g., Levine, 1983). Although authors disagree on the likeliness of solving the hard problem of consciousness (with perhaps a few exceptions), most everyone agrees it has yet to be solved and acknowledge that it poses uniquely challenging difficulties.

Schooler's Note on the Hard Problem of Consciousness[1]

One way of illustrating how the hard problem of consciousness differs from all other problems in science is to consider the following thought experiment. Imagine that an elite group of scientists who have received every imaginable recognition and accolade told you that they had solved the hard problem of consciousness and had developed a technique that definitively discerned what possessed subjective experience and what did not. These scientists use their pioneering innovation on you and conclude that you do not in fact have consciousness…you just think you do. Would you accept their conclusion? You surely would accept any other conclusion such a group of scientists might offer. But in this case, I hazard, you would be absolutely certain they had done something wrong. Ultimately, unlike anything else one knows, when it comes to the existence of one's own subjective experience, one's own first person perspective a priori trumps even the most authoritatively imaginable scientific evidence.

Dennett (1995) has argued that there is no hard problem of consciousness because there is nothing particularly special about the first-person perspective. If I understand him (and I have to confess, despite my best efforts, I am far from sure that I do), his position is that there is ultimately nothing privileged about the first-person perspective that cannot be comparably extracted from the third-person perspective. Or put another way, any information that we believe we possess exclusively from the first-person perspective is ultimately suspect. Although I am quite sympathetic to the view that introspective evidence can be questionable (Schooler & Dougal, 1999; Schooler & Fiore, 1997; Schooler & Schreiber, 2004; Wilson & Schooler, 1991), it *seems to me there is at least one thing that we can only know from the first-person perspective, and that is that we are conscious. In a prior writing, we (Schooler & Schreiber, 2004) interpreted Dennett's dismissal of the significance of qualia as indicating that conscious experience is itself an illusion. Although we were in good company in this reading of his "qualia disqualified" chapter in Consciousness Explained (1991; see Searle, 1997). I have learned both from other writings of his (e.g., Dennett, 1997) and from personal communications that he does not deny the existence of subjective experience. The question that I still have not resolved, however, is how he knows that subjective experience exists, if it is not on the basis of his own first-person perspective.*

Ultimately, it seems the unique knowledge afforded by subjective experience is precisely that which cannot be communicated to anyone else. Thus, the only way to argue for such knowledge is to appeal to individuals' own first-person experiences. It is tempting to suggest that those who deny the importance of qualia are somehow lacking the unique qualities of subjective experience that seem so self-evident to others of us. Having watched Dennett enjoy a fine meal with as much relish as anyone, I rather doubt this view, but the fact is I can't be sure…and therein lies the hard problem of consciousness.

THE EASY PROBLEMS OF FREE WILL

In our view, the distinction between the hard and easy problem of consciousness has important parallels to the problems associated with free will. Like consciousness, there are some questions about free will that seem to be straightforwardly addressable within the context of existing cognitive frameworks. As will be seen, issues such as the timing of the experience of will, the situations under which the experience of will arises, and situations in which these experiences are misguided can all be naturally studied and naturally understood within existing psychological/mechanistic frameworks. The hard problem of free will arises in reconciling the subjective experience of genuine personal choice with scientific claims that all actions are the necessary consequence of preexisting causes. Just as there seems to be no place to put subjective experience into the material formula, so, too, there seems no place for genuine choice. In the following discussion, we first consider some of the important work that has been done in addressing the easy problems of free will. Importantly, the work included in this category is far from trivial, and indeed in many cases has offered major insights into understanding the concomitants of an experience of will. Moreover, in contrast to the studies addressing the easy problem of consciousness, studies addressing the easy problems of free will do offer some constraints on the hard problem of free will, in the sense that they set some important boundary conditions on what people might potentially have control over. Nevertheless, as will be argued, such studies fail to put the proverbial nail in the coffin of free will, as they neither establish nor disconfirm whether conscious thought can ever be responsible for action.

Serious scientific study of free will and conscious volition can be said to have begun with Libet's experiments on the timing of conscious choice. In order to locate the temporal position of the conscious willing of an action in the chain of events that led up to the performance of a voluntary action, Libet and his colleagues (1983) designed an apparatus to measure self reports of volition and compared them to the readiness potential—an EEG measured indicator of when the neural processes that give rise to the action begin. Libet and colleagues found that what they identified to be the volitional impulse to begin the action occurred around 350 ms *after* the readiness potential had begun. Without temporal primacy, it is very difficult to attribute causation to the volitional impulse. Further research (Libet, 1985, 1999) went on to demonstrate that a cognitive mechanism craftily reversed these temporal positions in the conscious mind. The actual time of the conscious impulse was "backward referred" about half a second so that it seemed to have occurred before the action began. What people were becoming conscious of was a false memory of having decided to act, rather than an actual desire.

Almost all of the works involved in the recent deluge of anti–free will arguments have referenced this study despite a steady stream of criticisms questioning the actual meanings of the time delay and backward referral (Gomes, 1999; Klein, 2002), the self-reported conscious desire (Dennett, 1995) and the "freedom" of the involved action (Levy & Byne, 2004). Nevertheless, Libet's experiments have stood the test of time and become the scientific spine of the anti–free will movement. Interestingly, however, Libet's own interpretation is considerably different from this prevailing view.

However Libet's work has been interpreted since, he refrained from concluding that his results could definitively inform the hard problem of free will. Instead of abandoning the possibility that free will exists, Libet has taken an approach less reliant on traditional materialist methods of explanation. He proposed that free will exists in the reduced capacity of a selection process—what he calls the veto clause. Although actions are generated unconsciously, he suggests there is a conscious control function that gives the option to allow or prevent that action from coming to fruition. Libet explains that because this veto power would represent the genuine type of conscious free will that can exist only unconnected to any physically determining forces (an "unmoved mover" in Clark's [1999] language), no directly observable material evidence should be expected. Instead, as this is a genuinely conscious and subjective power, science should look to introspection and the self-reports of his subjects to support the existence of the phenomenon (this approach is discussed further in the final section). It needs to be recognized that the veto clause alone cannot be taken as serious evidence informing the hard problem. Instead of addressing how the mental could affect the physical, Libet assumes it does and then creates a mechanism that could exploit this assumption. Progress into the hard problem would require a discussion of the mechanisms by which the subjective power of the veto clause actually affects the physical neurons of the brain, how it actually moves the meat. Until this happens, the veto clause remains an interesting mechanism that evades the actual question that Libet wants to be addressing.

Shariff and Peterson (2005) have made another attempt to reconcile the time lag with at least an indirect form of conscious control. This "close-enough theory of free will" suggests that people's actual actions are unconsciously initiated by well-learned schemata that link together object perception and associated motor actions. The conscious control over such actions is relegated to the switching of attention in object perception. By choosing to pay attention to various salient features of your environment, you elicit the relevant action schemas associated with that stimulus. Which schema will be initiated is probabilistically biased through a modulation of activation strength—so perceiving a glass of water as a "container of drinkable fluid" makes it highly likely that the motor schema associated with "grabbing and drinking from said container" is elicited. Though this is not direct free will, the authors suggest that it is something close enough. Their key insight is that instead of having complete control over one's thoughts and behavior, people have some sort of detached and lower resolution control. An apt analogy may be to consider sailors who have no control over the wind and waves when sailing, nevertheless, they can set their sails and plot a general course.

Unlike Libet's veto clause, Shariff and Peterson's theory is embedded within and built out of a collection of converging clinical, cognitive, and theoretical support. Research on the peculiar neurological disorder of utilization behavior is synthesized with evidence from cognitive neuroscience as to how perceptual-motor "action macros" are construed and unconsciously run. A revised version of Gibson's (1979) direct perception is interpreted in the context of modern theories on attention. Finally, the authors suggest how this close-enough theory can be reconciled with the timing issues introduced by Libet's original data. But like Libet, the Shariff and Peterson theory does not tackle how conscious free will could exist. Instead, the theory proposes only how free will could work, with its existence assumed. Until this theory explains how conscious subjective selves can affect the redirection of attention in object perception, it does not address the central problem, or, in our terms, the hard problem.

Wegner's (2002) argument faces the same limitations in addressing the hard problem of free will, but from the other camp. In *The Illusion of Conscious Will* (2002), he explores many examples of cases in which will is perceived, but is demonstrably uninvolved in action. Cases such as hypnosis, alien hand syndrome, and those involving direct brain stimulation suggest to Wegner that the neural systems governing the experience of will may be quite detached from the mechanisms responsible for the actions themselves. Moreover, Wegner details people's tendency and perhaps motivation to perceive causal ownership to actions.

Wegner's explanations may be *necessary* to explain how a lack of conscious free will can be reconciled with the subjective experience of such volition, but it is not *sufficient* to dismiss such volition as an illusion. The existence of conscious free will is still fully compatible with the various illusions and mechanisms that Wegner describes. Like Libet, Wegner acknowledges the scope of his argument:

> Questions of whether thought actually does cause action, for example, have been left in peace, and the issue of the role of consciousness in the causation of action has been ignored as well. This is because the focus of this theory is the experience of conscious will, not the operation of the will.…This theory is mute on whether thought does cause action. (Wegner, 2005, p. 32)

As has been demonstrated, neither Libet's data, nor the theories put forward by Libet and Shariff and Peterson, nor the mechanisms identified by Wegner address the core of the free will problem—the hard problem of free will. To use Wegner's words, they tiptoe "quietly around the big sleeping problem of real mental causation" (2005, p. 32). Each makes progress on surrounding issues, often in conflicting directions, but with regard to the hard problem, none does more than push it further back—shrinking the black box, but never opening it. Before we turn to some brave forays into that problem, it is worth looking at one more undertaking that could be considered an "easy problem" of free will.

Bargh and his collaborators have assembled a wide-ranging body of literature that demonstrates the extent to which people's actions, emotions, and beliefs can be shaped unconsciously (see, e.g., Bargh & Chartrand, 1999). In one study, the researchers demonstrated that the mere mention of words relating to stereotypes of the elderly, such as "Florida" and "wrinkle," caused a measurable and significant change in their walking speed. In another experiment, subjects primed with rudeness, as opposed to politeness, were more than four times as likely to interrupt a conversation between the experimenters (Bargh, Chen, & Burrows, 1996). A third showed that those subliminally primed with African American faces were more likely to behave aggressively in social interactions (an African American stereotype). Subsequent research has shown that unconscious priming can dramatically influence goal pursuit (Bargh, 1989), moral behavior (Shariff & Norenzayan, 2007), self-sufficient behaviors (Vohs, Mead, & Goode, 2006) and, remarkably, the amount of free choice people think they had in a task (Seguin & Pelletier, 2000). In all of these studies, the subjects remained unaware of the priming, claiming full ownership over their actions, and sometimes vigorously disagreeing with the experimenter when informed of how they had been manipulated.

These are important studies. They compellingly illustrate not only the massive capabilities of the unconscious mind but, combined with Wegner's work, the disconnect between the feeling of free will and free will itself. People are *easily* duped. At least some of what is perceived as spontaneous and endogenously generated choice is due to unnoticed environmental suggestion. There are, however, limits to what conclusions should be drawn from this research. Studying circumstances in which free will necessarily does not occur is not the same thing as demonstrating that free will can never occur. In many of the studies run by Bargh and colleagues, the standard procedure involves using an unconscious priming technique to observe the effects it has on some aspect of behavior. In each case, the participants' responses are, to a significant extent, the product of the experimenter's suggestion. Although keeping these suggestions outside of awareness is, of course, necessary for the purposes of the study, doing so restricts the opportunity to observe free will in the hard sense, or even the softer sense of a freedom of action. The priming technique is coercive and mere steps away from hypnosis. Using this research to directly inform the understanding of free will would be like using black-and-white photographs to study color blindness.

Moreover, like with visual illusions, illusions of will illustrate the manner in which subjective experience can be tricked, and thus offer important insights into sources of slippage in the system. However, to suggest that illusions of will demonstrate that the experience of will *never* has any bearing on what a person does would be akin to arguing that the existence of visual illusions demonstrates that there is never a correspondence between perception and external reality. Illusions of will importantly constrain the relationship between subjective experience and action, but they do not eliminate the possibility of such a relationship, and thus they leave the hard problem of free will intact.

THE HARD PROBLEM OF FREE WILL

The hard problem of free will boils down to whether the subjective experience of volition influences action, and if so, how? Like the hard problem of consciousness it involves understanding the mechanism underlying the interface between the mental and the physical. However, unlike consciousness, the existence of which is demonstrated by the very experience of subjectivity, the experience of free will does not in itself prove that it exists. Free will *could* be an illusion. This highlights a difference, then, between philosophical zombies and robots. Whereas a zombie is construed as a being that acts identically to a conscious human but lacks conscious experience, a robot, as we describe it, would still have those experiences but they wouldn't do anything. The robot would be driven entirely by its internal and unconscious machinery. Its consciousness would be an epiphenomenon—its perceived efficacy, an illusion. The most important difference is that people personally know, via introspection, that they are not zombies but they cannot know whether they are robots.

The similarities and differences between the hard problems of consciousness and free will require a reexamination of which methods of study can be trusted.

What role should introspective data play in understanding conscious free will? Libet, for instance, relies heavily on such introspective data to support the existence of his veto clause. Noting that consciousness, residing in a totally different ontological realm, is observable only subjectively, he explains that a mechanism involved in true conscious free will would exist in the same way—unapparent to direct objective observation but readily apparent phenomenologically. The veto mechanism fits this description, and as there is no clear evidence contrary to its existence; Libet argues on introspective grounds that it must exist.

Unsurprisingly, not everyone agrees. Clark (1999) lambastes such mentalist theories. "Feelings or intuitions," he says, "never count as self-evident proof of anything" (p.286). But this is not wholly true. We reviewed one case, perhaps the only one (but certainly an important and relevant one), in which feelings and intuitions *do* count as proof. If the feeling of consciousness did not matter, if it contributed nothing toward proving that it exists, then science wouldn't have any of these problems—easy, hard, whatever. This is where the role of introspective data, which is easily dismissed by the materialists, becomes thornier. Were one to take the line of reasoning that the materialists use to explain the infeasibility of conscious will and apply it to consciousness itself, one would have to conclude that there is no such thing as subjective experience at all. Were one to use, for consciousness, the line of reasoning that the materialists are using to explain the infeasibility of conscious will, one would quickly conclude that there was no such thing as subjective experience itself. But the prima facie introspective self-evidence of the existence of consciousness directly contradicts materialist arguments that it cannot exist. The fact that people's feelings and intuitions conflict with the materialist argument at least casts suspicion on the comprehensiveness of materialism as it is currently understood. Consciousness manages to exist, in spite of the fact that from a material standpoint it should not. It is an anomaly that has yet to be satisfactorily accounted for. Hence, when considering conscious free will—disparaged by all the same argument that should make consciousness go away—one wonders if there isn't room for another anomaly in this otherwise physical world.

Of course, speculating that free will is a plausible second anomaly in an otherwise materialist world must be done with great caution. Just because consciousness seems to trump materialist arguments on introspective grounds does not mean the floodgates should be opened for other nonmaterial constructs. Whereas each one of us would be willing to stake our very lives on the introspective certainty that we are conscious, perhaps none of us would be prepared to do the same for free will. Nevertheless, the two apparent phenomena seem to emerge from the same mystery—a misunderstood relationship between physical matter and subjective sensation. Given this connection and given the introspective power of the experience of personal agency, it should at least lead scientists to entertain the possibility that free will may similarly exist despite its material complications.

Ultimately, the landscape of the hard problem of free will may become clearer than it is now once there is a reasonably robust solution to the original hard problem of consciousness. Unfortunately, such a solution has been far from forthcoming—with some going so far as to suggest that such a solution is theoretically impossible to be grasped by the human brain (McGinn, 1999). Whether or not this position is true, it is worth pressing on and examining the way different theorists have at least attempted to address the problem. Although we refrain from articulating a single approach, we encourage the reader to keep an open mind as to the faults and strengths of the following positions, which we believe to be promising directions.

There are three broad camps in the battle for free will—hard determinist, compatiblist, and libertarian—and membership to these groups by and large falls along materialist-mentalist party lines.

Hard Determinism

The determinists are the most likely to dismiss free will as an illusion and reduce human beings to robots. They take the reductionist position suggested by Crick's quotation at the beginning of this chapter as their starting point. In the causal chain of behavior, they say, there is no room for anything beyond the story told by the physicists and neuroscientists. All behaviors and cognitions are initiated by the underlying "machinery" of the brain—whereas consciousness is another unidirectional product of these neural processes. It has no reciprocal effects on cognition, making conscious free will not only an illusion, but an impossibility. The brain simply does not work that way.

A quotation by Samuel Johnson that determinists are fond of using reads "all science is against free will, all experience is for it." This is absolutely true. With the exception of some research on the easy problems of free will (which *should*, as we've discussed, be excepted), there has been no evidence from traditional science that has provided a compelling case that free will in the sense of conscious agency exists. Experience certainly has, but as the determinists will be quick to point out, when pushed, experience alone has historically never been a good indicator of anything. Whether consciousness,and subsequently conscious free will,break this historical mold and should be given exceptional status in the realm of science depends on just how special and unique a phenomenon one thinks consciousness is, which isa matter of no small disagreement between researchers in the field.

Finally, it is important to note that when it comes to consciousness and free will, impotent does not mean the same thing as useless. This is a crucial distinction that is often lost in the imprecision of our lexicon. Although conscious free will may not exist, this does not mean that the *idea* of conscious free will does not exist. The belief in free will exists in the brain, and, as the Vohs and Schooler data show, the belief in free will *is* involved in the causal chain of behavior. If it is an illusion, the illusion is an efficacious one and insofar as it prevents moral laxity, it is a *useful* one. It keeps people feeling morally responsible as agents, although, in truth, they may not be. There may be a time, if the hard determinists are proven are right, that people will choose to maintain a belief in free will that they know to be disingenuous—acting as if free will still existed and deceiving themselves into good behavior. But we are getting ahead of ourselves; the determinist position has not been proven, nor has it been disproven. It has, however, been challenged extensively, and there other positions merit consideration.

Compatibilism

The term compatiblism is used generally and originally with reference to the debate within philosophy about free will in the context of an allegedly physically determined universe—a debate that we have stayed away from. Therein, it referred to the position that free will could coexist with—be *compatible* with—a deterministic view of the universe. We modify this meaning only slightly to refer to psychological compatibilism that is, as one would expect, the position that attempts to reconcile free will with the apparent impotence of consciousness. Though it is a broad category, most psychological compatibilists tend to eschew appeals to material dualism without completely dumbing humans down to robots. They maintain, in one way or another, that despite the deterministic laws people must follow, they nonetheless have genuine options and opportunities for action.

Certain compatibilist interpretations are, no doubt, less compelling than others. For instance, some seem to be saying that so long as people do not feel as though their choices are psychologically or physically constrained, then they enjoy free will (Strawson, 1998). In the light of both Bargh's (e.g., Bargh & Chartrand, 1999) and Wegner's (2002) research, this position falls prey to fatal weaknesses. The feeling of free will, as Bargh and Wegner have shown, is an easily fooled indicator of true free will—even in the sense that these compatibilists mean. Bargh's subjects in numerous experiments were undoubtedly psychologically compelled to behave in the way they did, but felt throughout that they were exercising their free choice in all their endeavors. Feeling, we must repeat, is not enough. It is an idea worth considering that the *belief* in free will may be the only "variety of free will" worth wanting, but this is a separate debate. Ontologically, this form of compatibilism is left wanting.

A more sophisticated compatibilist approach stems from Velmans' *dual-aspect* approach. Dual-aspect theory is an attempt to frame the hard problem of consciousness by employing a softer form of Cartesian dualism. Velmans and others square a materialist understanding of the world with the apparent differences between third-person neurons and first-person experience by arguing that neurons and experience can be made of the same materials, but still have different identities. Velmans (2002), specifically, explains that although the two identities share the same "informational structure," they are ontologically different. This type of theory, which shows first- and third-person experience to be *equivalent*, but not *identical*, is known as identity dualism.

Relating identity dualism to free will provides a potentially promising theory of volition. Clark (1999) outlines such a position, an informed synthesis of both Gomes (1999) and Claxton (1999) from the same volume. Moving from the understanding that the subjective mind *is* brain-based activity, he explains that conscious desire doesn't spawn or lead to neural processes any more than neuronal activity spawns or leads to conscious experience. There is no two-step process. Instead, the two are equivalent things; the experience of conscious free will is the first-person perspective of the neural correlates of choosing. This means that although there is no nonphysical, conscious self-homunculus that sets things in motion, there is still an "I"—the holistic compendium of your brain processes over which people can and should take ownership. It is this holistic self with which people should identify. At the risk of confusing the issue, one could say there is no prime mover, but rather the movement is the mover.

In some ways, being compelled by compatibilist approaches like this one leaves a person with the feeling that he or she had a run-in with a known pickpocket: One should probably do a quick pat-down to see if anything important was taken. When considering the Velmans-Clark-Gomes-Claxton approach, it is worth asking what aspects of free will are maintained, what are lost, and what, if any, are gained. Ultimately, people maintain most of the types of free will (worth wanting) that they had when assuming that they possessed a nonphysical, conscious controller. They remain responsible for their actions; and they remain in charge. What they want still happens.

Nonetheless, there are subtle but deceivingly important differences between the folk understanding of free will and this current reconceptualization. In the former position, one's "I" is understood to refer to "the me that does thing thinking" and this self is credited with being the one that consciously controls one's actions. The new approach dissolves the conscious self into the larger "I" and "the me that does the thinking" is embedded within the whole brain. "I" still control my actions, but the "I" is reconceived to be the coalition of my brain processes. And although I am still fully in control of my actions—although my will is still free—the nature of conscious free will changes. Instead of saying that people have conscious free will, we must instead say that people are conscious *of* their free will. Instead of saying that my consciousness (me) is making the decisions, we need to say that I am conscious of the parts of my brain (still me) that are making the decisions. Instead of saying the "I" moves the machinery of my brain, "I" *am* the machinery of my brain, and "I" consequently move myself.

Although this distinction may initially seem to be just semantic—rhetorical sleight of hand—closer inspection yields important differences, both philosophical and pragmatic. Most importantly, it seems that one loses the ultimate sense of free will that Kane (1996), and many others, believe worth wanting. Giving up control from the conscious, agentic self to the holistic "you are your brain" controller means just that. The "I" that one usually identifies with—one's conscious self—loses its claim on causation. Insofar as people continue to identify with the former conscious agent, they do not have free will in the way that Kane defines it: "*the power of agents to be the ultimate creators (or originators) and sustainers of their own ends and purposes*" (p. 4). To say, as the proponents of the new position will, that people maintain this free will by expanding their conception of themselves as agents to their holistic brain is fine and by a certain interpretation true, but it introduces a host of its own problems.[2]

Libertarianism

Libertarians, like the hard determinists, are incompatibilists. However, unlike the determinists, they believe that human beings still maintain a sense of ultimate agency and control over their own will.

The libertarian position is both the one that is most aligned with traditional, intuitive understanding of free will and the one that is most often, and most vociferously, maligned. So out of vogue is this position with the majority of consciousness researchers that Wegner, with only some irony, divided the field into "robogeeks" (those who espouse something approaching the determinist position) and "bad scientists" (those who stick to their intuitive beliefs despite the wealth of conflicting evidence). And the libertarians have been criticized, with some justification, as resorting to increasingly "panicky metaphysics" to maintain their belief in ultimate agency. By clinging to traditional conceptualizations of free will, they have been accused of resisting the progress of science and being unwilling or unable to update to new paradigms. This may be true to some extent. But this argument could also be made in the opposite direction: The assumptions involved in scientific examination may themselves be in error. The existence of consciousness and the apparent existence of conscious volition may be examples of anomalies that indicate the limits of our current investigative paradigm. And those researchers who are strictly abiding by the established materialist modes of investigation may be the ones who are being overly rigid, trying vainly to cram ever more complex phenomena into inadequate methods of explanation. It is perhaps not the traditional understanding of free will that is in error but, rather, the traditional understanding of how to do science.

Libertarians point, not unreasonably, to increasingly bizarre discoveries in other scientific endeavors such as quantum indeterminacy, discoveries that have forced scientists to reevaluate their preexisting assumptions. Coupled with the relative ignorance of the brain sciences, libertarians insist that there is ample room left for the claims of determinists of any persuasion to be overturned.

These claims are all certainly true, but it is this attitude that proves to be the greatest weakness of the libertarian position. From the start, the libertarians have been playing defense—trying to maintain a status quo position against a tide of criticisms. The assumption has been that the burden of proof was on those attempting to dispel the existence of free will. This has not been enough. For the libertarian position to be taken seriously, its adherents need to demonstrate its viability, and not just its possibility.

Schooler's Reflections on the Hard Problem of Free Will

In principle, the compatibilist perspective is ideal as it allows us to have our material cake while freely choosing to eat it, too. The problem is that I, for one, simply cannot get my head around compatibilism. I understand that determinism does not rule out the opportunity to make deliberate decisions that are free from coercion, I acknowledge that I can redescribe the control of my actions as being completed by my brain as opposed to my mind, and I admire Dennett's (2003) argument that the evolution of culture and the human brain's capacity for rational analysis has enabled individuals to make reasoned decisions. I just don't understand the following: If any noncoerced reasoned decision that I am about to make must necessarily be carried out in a specific manner based on a preexisting causal chain, then how can I be free to choose otherwise? And if I really have no option but to do exactly what I end up doing, how can it be said that my choice was free? On this point, I find myself in agreement with Greene and Cohen (2004) who argue, "…contrary to legal and philosophical orthodoxy, determinism really does threaten free will and responsibility as we intuitively understand them." (p.1780)

Given the indefensibility of compatibilism, Greene and Cohen (2004) go on to argue that neuroscientific evidence supports the hard determinist view that "free will as we ordinarily understand it is an illusion generated by our cognitive architecture." (p.1784) They, like many others, dismiss libertarianism as "panicky metaphysics" noting, "…there is not a shred of scientific evidence to support the existence of causally effective processes in the mind or brain that violate the laws of physics." (p. 1777). Importantly, however, an absence of evidence is not necessarily evidence of an absence. For example, there is no scientific evidence indicating the existence of other universes; nevertheless, many rational individuals have postulated that alternative universes might exist (Greene, 2003). The problem is that just as there is no evidence that alternative universes do exist, there is also no evidence that they don't. Similarly, although we have made important strides in understanding human behavior, we are still very far from perfectly predicting human action. Given the current indeterminacy of human behavior, there is still room for additional causal sources. Thus although there may be no scientific evidence for the existence of free will, there is also no scientific evidence of its absence.

Moreover, and here I go out even further on the limb, there actually is some evidence that consciousness can influence the outcome of physical events in a manner that, if true, would appear to violate at least our current understanding of physics. Specifically, a recent review (Bösch, Steinkamp, & Boller, 2006) in the prestigious journal Psychological Bulletin reported the results of a meta-analysis reviewing 380 studies that examined whether "random number generator output correlated with human intentions" (e.g., whether participants' efforts to will a random number generator to produce odd numbers increased the likelihood of it doing so). The result of this analysis revealed a "significant but very small overall effect size." (p.497) The authors cautiously concluded that "publication bias appears to be the easiest and most encompassing explanation for the primary findings of the meta-analysis." However, they also conceded that this explanation would require a rather large number (n = 1500) of unpublished studies. Moreover, they acknowledged that the alternative possibility that human intention was influencing random events could not be ruled out, noting, "The effect in general, even if incredibly small, is of great fundamental importance—if genuine." (p. 517)

Clearly, remarkable claims require remarkable evidence, and this report, though striking, is unlikely to persuade many that they should accept the claim that human consciousness can influence physical events. At the same time, it does constitute at least a "shred of evidence" that causally effective processes in the mind can influence physical processes. And if conscious will is capable of influencing random number generators at a distance, then surely it should be capable of influencing the far more proximal behaviors of the brain.

It is important to emphasize that even if mental intentions can influence physical events, it does not necessarily follow that free will exists. The human intentions that potentially influence random number generator could themselves be the necessary consequence of causal chains. My argument is simply that given the existence of at least a shred of evidence for causal effects of human consciousness, and given the degree to which the sources of human thought and action are still not understood, its seems premature to conclude that scientific evidence definitively rules out the possibility of a genuine impact of conscious will.

Ultimately, the viability of the libertarian perspective will depend on the generation of accounts that are both phenomenologically compelling and scientifically tractable. In this regard, I am sympathetic to the view that quantum physics may offer a possible opening for genuine free will. There are a number of characteristics of quantum physics that offer some hope for salvaging of free will. First, quantum physics demonstrates that causal determinism is not, as is often suggested, a necessary aspect of the universe that is fundamentally violated by the notion of free will. If, as quantum indeterminacy suggests, the future is not written in stone, then different actions could lead to different futures. It is often pointed out that quantum indeterminacy offers little solace for libertarians, because having one's choices influenced by a combination of deterministic forces and some random quantum element still leaves no room for the conscious chooser. As Greene and Cohen (2004) observe, "If it *turns out that your ordering soup is completely determined by the laws of physics, the state of the universe 10,000 years ago, and the outcomes of myriad subatomic coin flips, your appetizer is no more freely chosen than before" (p. 1777). However, this presumes that the indeterminacy observed at the quantum level is entirely random. If, as the studies on random number generators mentioned above hint, the mind can influence the outcome of random events, then indeterminacy associated with quantum effects might appear random only because we have failed to assess the causal impact of consciousness.*

A second important feature of quantum indeterminacy is its peculiar relationship with observation. Although accounts of the role of observation in quantum outcomes remains an area of considerable dispute, many reasonable scientists have argued that the way in which we observe quantum events influences how they unfold (for a fascinating and highly readable account of the challenges of understanding the relationship between physics and consciousness written by noted physicists, see Rosenblum & Kutner, 2005). If consciousness can influence external events, then why couldn't it similarly influence internal mental events? Admittedly, at present, it is not clear how quantum indeterminacy could operate at the macro level of warm brains. Although some have speculated about possible mechanisms (e.g., Hameroff & Penrose, 1995), others have highlighted the various reasons that such accounts are implausible (Koch & Hepp, 2006).

Such disputes lead to the final important lesson of quantum physics, which is that although its properties can be described with remarkable precision, the explanation for why it interacts with observation in the way that it does remains a central mystery. Given the further inability of science to account for consciousness more generally, it seems that we simply do not understand the nature of either physics or consciousness enough to know with assuredness what their relationship might be. Were subjectivity and volition modest aspects of our existence, then it would seem entirely unreasonable to suggest that some unforeseen revolution in scientific understanding might offer them a greater place. But to the contrary, the experience of being and doing are arguably the two most essential aspects of our day-to-day lives. Although science has made incredible advances, has it really progressed to the point at which it can reasonably ask us to disregard the defining aspects of our existence?

THE CONSEQUENCES OF A BELIEF IN FREE WILL

Although the above approaches to the hard problem of free will ultimately be judged on their scientific and philosophical merits, there are undoubtedly other factors affecting the attractiveness of each position. The libertarian position, for instance, has been vigorously defended in large part because of phenomenological factors. There is the intuitive appeal of the mind working the way one feels it does. There is also a sentimental attachment to the belief that as human beings, we are each prime movers. Perhaps as a consequence, there is a visceral repulsion to the determinist doctrine of fully automated robots. Agency and responsibility are tied to meaning, and being robbed of agency likewise robs one of meaning (Heine, Proulx, & Vohs, 2006). The compatibilists, meanwhile, seem to be resigning themselves to the indefensibility of the libertarian position, but trying to salvage as much meaning as they can through complex arguments.

Beyond the psychological palatability afforded to each position, there is also the moral component. The question of free will has always been tied to the moral implications of the argument (Pereboom, 1997). Generally, the predictions made about the consequences of an eroding sense of free will differ systematically with the position they are trying to advance.

The libertarians tend to present morally bankrupt dystopias as the inevitable consequence of abandoning a belief in ultimate agency and responsibility. The argument, an old one, can be summed up as follows: Without belief in causally responsible agents who could have done otherwise, there is nothing to blame or praise and therefore no way to sanction moral behavior. Without such sanctions, without the social and personal prescriptions that come with recognizing what is worth praise or blame, and without the personal dignity that comes with agency, people will be reduced to selfish beings without a moral compass. This position has been recently articulated by Bennett (1998) and Goodwin (1998), the latter of whom believes it to be held by the lay majority.

Determinists and compatibilists present rosier views. Greene and Cohen (2004) predict that as people become more accepting of compatibilism in the justice system, they will move from a punitive system demanding retribution to a more humane system more concerned with the consequences of punishment. Clark (1999) agrees, adding that by replacing the latently metaphysical beliefs that most laypeople hold with "a thoroughly naturalistic conception of the self and its choices" (p. 17), society may experience a less punitive culture more focused on the exogenous causes of individual ills, whereas individuals, themselves, will benefit from a lack of self-consciousness and personal blame. This optimistic outlook also includes less wealth hording as ambition softens and, with a shift in societal understanding of freedom, a "more responsible use of such freedom" (p. 18).

Unfortunately, empirical investigation from our lab suggests otherwise. People who are disabused of the illusion of agentic control seem to, at least temporarily, abandon their moral code. Two experiments that manipulated participants' belief in free will show that when people come to believe that the idea of free will is untenable, they behave amorally, for which we found evidence in the realm of cheating (Vohs & Schooler, 2008).

In the first experiment, participants were induced to believe or not believe in free will via reading an essay written by Nobel laureate (given for the codiscovery of DNA) Francis Crick. One chapter of Crick's book, *The Astonishing Hypothesis* (1994), claimed that rational, thinking people (such as scientists) long have denounced the idea of free will, noting that it is instead a byproduct of the human mind. Participants who read this essay were in the anti–free will condition, whereas control condition participants read from another chapter in this same book on consciousness, which did not contain references to free will. Afterward, participants were seated in front of a computer that displayed, one by one, multicomponent mathematical problems that participants were supposed to calculate in their heads. We told them that due to a programming error, the computer had a glitch that allowed the answer to be shown after a short period of time. Participants were told that they could, however, stop the answer from being shown by pressing a certain key on the keyboard. This situation, then, gave participants the opportunity to cheat, but also gave them a simple way to avoid it—a slight movement of the hand. Would convincing participants not to believe in free will alter whether they let themselves cheat?

As expected, it did. Participants who were disabused of the idea of free will cheated more by letting the answer appear relative to participants who read an essay that did not speak to the existence of free will (Vohs & Schooler, 2008). The total number of math problems that participants saw was 20, and the anti–free will group let themselves cheat on almost 12 problems, whereas the control group let themselves cheat 9 times.

Some readers may be wondering whether the Crick essay effectively changed participants' beliefs about free will, or whether the effects were due to some other cause. To test for changes in cognitions, we included a scale that measured belief in free will (Paulhus & Margesson, 1994). As expected, participants in the anti–free will condition reported lower scores on the free will subscale compared to participants in the control condition. Moreover, belief in free will scores correlated significantly (and negatively) with the propensity to cheat. In other words, after reading the anti–free will essay, compared to measures taken after reading an essay that was devoid of free will information, participants said they were unconvinced that free will exists and they let themselves cheat.

Although these results were promising, we recognized that there are alternate explanations of these results. It may be that although cheating behavior changed, participants were not acting immorally but rather behaving passively as a result of the anti–free will essay. Hence, we conducted another experiment to rule out this explanation, using a cheating behavior that required active participation to cheat. We also added a condition that enhanced beliefs in free will, which provided an opportunity to test the full effects of believing or denying the existence of free will.

In this second experiment (Vohs & Schooler, 2008), participants received one of three treatments. In one condition, we had participants read a series of statements designed to induce a feeling of determinism, which we believed would also have the effect of reducing free will. Sample statements included, "Ultimately, we are biological computers—designed by evolution, built through genetics, and programmed by the environment." Participants' task was to read each statement and think about it, and then when instructed they were to turn the page and read another statement. This task is modeled after the oft-used Velten mood induction task (Velten, 1968). In another condition, participants read statements that were designed to bolster beliefs in free will, such as "I am able to override the genetic and environmental factors that sometimes influence my behavior." A third group of participants read neutral statements.

The cheating opportunity was set up such that participants self-scored a cognitive test on which they were to be paid $1. Ostensibly because of an unexpected errand, the experimenter left the room and allowed participants to score their exam and then pay themselves for their performance on the test. We compared the money participants paid themselves, as a proxy for their claimed scores on the exam, to veridical scores from participants who took the exam and were not allowed to self-score. The research question was whether participants would give themselves differential amounts of money as a function of whether they had been induced to belief in free will, determinism, or whether their beliefs were left unchanged.

The results showed that after participants read statements that told them their actions were predetermined and therefore not under their control, they cheated more—as evidenced by more money taken in this condition compared to the control condition and the free will condition. Reading statements that bolstered participants' belief in free will did not affect cheating behavior, as these participants paid themselves as much money as did participants whose scores were known. Once again, we knew that participants' beliefs did change, as evidenced by changes in the Free Will and Determinism Scale (Paulhus & Margesson, 1994) as a function of condition.

Hence, telling people that free will does not exist or telling them that their behavior is caused by predetermined mechanisms outside of their control leads them to cheat more so than people who are not induced to change their beliefs about free will. Note, interestingly, that in the second experiment, the control condition and the free will bolstering condition were not significantly different from each other, which suggests that lay beliefs about free will are in line with ideas of ownership and authorship of one's own behavior (cf. Wegner, 2005).

These data lend empirical support to the suspicion that eliminating people's beliefs in free will may be accompanied by reduced ethical behavior. These data do not mean, of course, that scientists should put all the weight behind the libertarian position. Nor does it mean that society will, in the long term, never be able to morally adjust to compatiblist[3] or even determinist worldviews. Instead, these data suggest that there are real consequences to scientists' claims. There may come a time when the evidence will require society to rethink its conceptions of freedom and responsibility—as Dennett says, "institutions and practices based on obvious falsehoods are too brittle to trust" (2003, p. 290). But the key word in that statement is *obvious*. Nothing, yet, is obvious. Given the present lack of a solution to the hard problem of free will and given the societal repercussions of convincing people that they lack genuine control of their behavior, it seems that caution is warranted when making assertions regarding free will outside of the ivory tower.

NOTES

1. In several places in this chapter, we have agreed to break with the convention of writing co-authored papers with an exclusively unified voice. The metaphysical ramifications of free will and consciousness are particularly contentious, and one of us (Schooler) has controversial opinions on the matter that are not necessarily shared by the other authors. As a compromise, Schooler has inserted several brief reflections sections (set apart from the main text) into this chapter that should not be assumed to reflect the views of Shariff and Vohs.

2. Some of these problems are particularly evident in situations in which the system has broken down. Consider, for example, alien hand syndrome, wherein one's limb acquires a "mind of its own" often interfering with willed activities and even attempting to choke the "host" (Scepkowski & Cronin-Golomb, 2003). In this case, the sense of control that we are conscious over is overwhelmed by the control of another facet of the brain. The brain has divided itself into two competing agents. It would be hard to convince the patient to identify with his or her holistic brain when unconscious parts are rebelling against the conscious parts. In this case, it is hard to see the pragmatic difference between having alien hand syndrome, in which the hand is controlled by endogenous factors, and having a condition in which one's hand has actually come under the control of exogenous aliens. To call this free will is surely to strip the term of any meaning. Although this example and others like it (e.g., schizophrenic delusions of alien control or more commonplace diseases such as Parkinson's disease) represent clinical situations in which the mechanisms of will and ownership have gone awry, they are illuminating and worth considering.

It isn't only in clinical cases that such issues arise, however. We have to consider, for example, what identification with the holistic self means for the unconscious or nonconscious decisions that are made by our brain. What about cases of sleepwalking, or reflexes or decisions induced by Bargh's experiments? These issues demonstrate, again, that adopting compatibilist positions will require people to sacrifice at least some of the intuitive, libertarian position that most have grown comfortable with. Moreover, these sacrifices might not be immediately apparent or easy to understand.

3. One important future direction is to examine the impact on moral behavior of exposure to a compatibilist perspective. Dennett (this volume) has suggested a description of a compatibilist worldview that he argues might not only avoid inducing immoral behavior, but might actually facilitate the moral behavior of participants. Notably, however, his suggestion of how to characterize the compatibilist worldview does not explain compatibilism; it simply assures the reader

that philosophers have worked it out. Thus the potential efficacy of his suggested manipulation may simply take advantage of people's willingness to accept claims on the basis of authority.

REFERENCES

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462–79.

Bargh, J. A., Chen, M., & Burrows, L. (1996). The automaticity of social behavior: Direct effects of trait concept and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230–44.

Bennett, W. (1998). *Defining the human spirit: Historical perspectives and contemporary challenges.* Presented at Neuroscience and the Human Spirit, Washington, D.C. September 24–25.

Bösch, H., Steinkamp, F., & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators—a meta-analysis. *Psychological Bulletin,132*, 497–523.

Breer, P. (1989). *The spontaneous self: Viable alternatives to free will.* Cambridge, MA: Institute for Naturalistic Philosophy.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

Clark, T. W. (1999). Fear of mechanism: A compatibilist critique of the "The Volitional Brain." *Journal of Consciousness Studies, 6*(8–9), 279–93

Claxton, G. (1999). Whodunnit? Unpicking the "seems" of free will. *Journal of Consciousness Studies, 6*, 99–114.

Crick, F. (1994). *The astonishing hypothesis.* New York: Scribner's.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Co.

Dennett, D. C. (1995). Facing backwards on the problem of consciousness. *Journal of Consciousness Studies*, 3(1), 4–6.

Dennett, D. C. (1997). *Kinds of minds: Towards an understanding of consciousness.* New York: Basic Books.

Dennett, D. C. (2003). *Freedom evolves.* London: Allen Lane.

Gibson, J. J. (1979). *Ecological approach to visual perception*. Boston: Houghton Mifflin.

Gomes, G. (1999). Volition and the readiness potential. *Journal of Consciousness Studies, 6*, 59–76.

Goodwin, F. K. (1998, September 24–25). *Opening remarks.* Presented at Neuroscience and the Human Spirit, Washington, DC.

Greene, B. (2003). The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory. New York: W.W. Norton and Company.

Greene, J. D., & Cohen J. D. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B, 359,* 1775–85.

Hameroff, S. R., & Penrose, R. (1995). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Neural Network World, 5*(5), 793–804.

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). Meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review, 10*, 88–110.

Kane, R. (1996). *The significance of free will.* New York: Oxford University Press.

Klein, S. (2002). Libet's temporal anomalies: A reassessment of the data. *Consciousness and Cognition, 11*, 198–214.

Koch, C., & Hepp, K. (2006). Quantum mechanics in the brain. *Nature, 440*, 611–12.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*(4), 354–61.

Levy, N., & Byne, T. (2004). A will of one's own: Consciousness, control, and character. *International Journal of Law and Psychiatry, 27*, 459–70.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*, 529–66.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies, 6*(8–9), 47–57.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain, 106*, 623–42.

McEwan, I. (2001). *Atonement: A novel.* New York: Doubleday.

McGinn, C. (1999). *The mysterious flame: Conscious minds in a material world.* New York: Basic Books.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*(4):435–50.

Nagel, T. (1986). *The view from nowhere.* New York: Oxford University Press.

Paulhus, D. L., & Margesson, A. (1994). Free Will and Determinism (FAD) scale. Unpublished questionnaire, University of British Columbia, Vancouver, Canada.

Pereboom, D. (1997). *Free will.* Indianapolis, IN: Hackett.

Pinker, S. (2002). *The blank slate.* New York: Viking.

Rosenblum, B., & Kultner, F. (2006) Quantum Enigma: Physics Encounters Consciousness. New York: Oxford University Press.

Scepkowski, L. A., & Cronin-Golomb, A. (2003). The alien hand: Cases, categorizations, and anatomical correlates. *Behavioral Cognitive Neuroscience Review, 2*, 261–77.

Schooler, J. W. & Dougal, S. (1999). The symbiosis of subjective and experimental approaches to intuition. *Journal of Consciousness Studies, 6*, 280–87.

Schooler, J. W., & Fiore, S. M. (1997). Consciousness and the limits of language: You can't always say what you think or think what you say. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 241–57). Mahwah, NJ: Erlbaum.

Schooler, J. W., & Schreiber, C. A. (2004). Experience, meta-consciousness, and the paradox of introspection. *Journal of Consciousness Studies, 11*, 17–39.

Searle, J. (1997). *The mystery of consciousness.* New York: Review Press.

Seguin, C., & Pelletier, L. G. (2000). *Automatic activation of intrinsic and extrinsic motivation.* Manuscript submitted for publication, University of Ottawa.

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Supernatural agent concepts increase prosocial behavior in an anonymous economic game. *Psychological Science, 18*, 803–9.

Shariff, A. F., & Peterson, J. B. (2005). Anticipatory consciousness, Libet's veto and a close enough theory of free will. In R. Ellis & N. Newton (Eds.), *Consciousness and emotion: Agency, conscious choice and selective perception* (pp. 197–215). Amsterdam: John Benjamins.

Skinner, B. F. (1971). *Beyond freedom and dignity.* New York : Knopf.

Strawson, G. (1998). Free will. In E. Craig (Ed.), *Routledge encyclopedia of philosophy.* Retrieved August 11, 2005, from http://www.rep.routledge.com/article/V014SECT1.

Velmans, M. (2002). How could conscious experiences affect brains? *The Journal of Consciousness Studies, 9*(11), 3–29.

Velten, E. (1968). A laboratory task for induction of mood states.*Behaviour Research and Therapy, 6*, 473–82.

Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science, 314*, 1154–56.

Vohs, K. D. & Schooler, J. W. (in press). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19*, 49–54.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wegner, D. M. (2005). Who is the controller of controlled processes? In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 19–36). New York: Oxford University Press.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*(2), 181–92.

10 Free Will and the Control of Action

Henry L. Roediger, III
Michael K. Goode
Franklin M. Zaromb

Free will is not a topic that much occupies psychologists' thoughts or writings. Yes, we may reflect on the topic now and then in a desultory manner, but it hardly weighs heavily on the minds of modern researchers. The American Psychological Association, in collaboration with Oxford University Press, produced an eight-volume *Encyclopedia of Psychology* with many hundreds of entries. The volumes covered seemingly every topic under the psychological sun, running the gamut from *archetypes* and *altruism* near the beginning, to *voyeurism* and *xenophobia* toward the end. However, there was no entry on free will. In fact, *free will* did not even appear in the index as a term mentioned once, anywhere, in all eight volumes! (*Free association* was mentioned a number of times, though.)

Checking various other authoritative sources, we found much the same result. Occasionally an author would agree with the notion that research psychologists assume there is no such thing as free will. After all, why study behavior and look for regularities and laws if people can just do what they please any time they want, and in any situation? That's a good question, so most experimental psychologists probably think about free will for a few milliseconds, and then they go back to designing their next experiment. Psychologists prefer to leave the topic of free will to philosophers.

The editors of this book challenge psychologists to think hard about the topic, even if they have not done so since that introductory philosophy course many years ago. The challenge is perfectly appropriate, even if the answers in this volume may go largely unread by practicing psychologists (sorry to relate this news, if someone thought otherwise). Still, we agree with the premise that those of us who study human behavior should think hard about free will. Perhaps the time when most of us think about this topic is when an experiment completely bombs. When subjects behave nothing at all as our precious theories and hypotheses say they should behave in a particular situation, the confounded psychologist may then come to believe in free will.

Does it really matter if free will exists? Human behavior is incredibly, overwhelmingly, complex. Edward O. Wilson in *Consilience* (1998) commented that the social sciences "are inherently far more difficult than physics and chemistry, and as a result they, not physics and chemistry, should be called the hard sciences." The determinants are so many that even if humans exert free will now and again, how would we ever know? Let us review some determinants here, but we will hardly be exhaustive. Starting with our genetic roots, even if we go back merely to the year 1500 or thereabouts, each of us has over 2 million ancestors. The gigantic genetic stew that each of us represents since then—happenstance meetings and couplings over the last 500 years—has led us to become the unique human beings we are today. The human genome is just beginning to be understood, but modern molecular behavioral genetics (agreeing with decades of twin studies) indicate that a good portion of human characteristics and behavior is genetically determined (in some cases, such as eyesight) and in other cases at least genetically influenced (scores on intelligence tests). After conception, all sorts of prenatal factors (what the mother eats and drinks, her hormone levels, her health or illnesses, her ingestion of drugs, and so on) affect the child who is born. And that child is born into cultures and circumstances that vary incredibly across the world. The culture and society in which we develop determine the language we speak, the food we eat, the behaviors we learn as appropriate for particular circumstances, among many other things. Anthropologists have long noted the huge influence of culture on behavior, although psychologists have come to study this topic only in the past 10 or 15 years.

Of course, just sticking with the factors noted in the previous paragraph, most psychological studies do not routinely consider genetics (not yet; it's coming, though), nor do prenatal factors occupy many researchers. Most psychological studies occur in Western cultures and most researchers don't give a thought about, say, communicating instructions for an experiment through the dominant language and simply assuming that they will be understood. Yet even within a relatively homogeneous culture, people's experiences differ greatly. Our parents, grandparents, teachers and schools, religious traditions and instruction, communities, and peers all help shape us as we grow. So do the books we read and the movies and TV shows we see.

The myriad factors listed in the previous paragraphs only begin to scratch the surface of determinants of human behavior. No wonder human behavior, even of the simplest sorts, is so hard to predict. No wonder so many differences exist among people treated the same way in the same conditions of our experiments. Our "error variance" is due not just to measurement error, but also to the hundreds and thousands of differences among people in so many characteristics. Given these complexities, how could we know if free will does exist? Might it be only one more factor of difference in the bubbling cauldron of factors that help shape behavior?

FREE WILL: SOME PRELIMINARY THOUGHTS

We might profitably stop the chapter with the last sentence and let the reader move on. However, we do think that empirical psychology has several lines of research that might help shed light on issues surrounding free will. The astute reader will note that we have up until now dodged defining the main term of interest. We will borrow for our purposes the definition provided by Wikipedia ("Free Will," 2006): "The problem of free will is the problem of whether human beings exercise control over their own actions and decisions." This definition works for us, because psychologists have studied the topic of how humans control their own behavior. If we convert the problem of free will to the issue of control of behavior, psychologists may contribute to the conversation even if they still dodge the central issue in their studies—does free will exist?

Several research traditions are of interest in the study of the control of behavior, although we focus primarily on four: the response-choice paradigm developed by Benjamin Libet, the stop-signal paradigm developed by Gordon Logan and his colleagues, the process-dissociation procedure developed by Larry Jacoby and colleagues, and the free and forced report procedure developed by Asher Koriat and Morris Goldsmith. These paradigms are concerned with the issue of conscious control of behavior. Although the study of the voluntary control of behavior should not be equated with the problem of free will, at least its study may shed light on important questions. We review the paradigms and some main findings in the next two sections of the chapter and then discuss implications in a third section. The chapter ends with some general thoughts and conclusions.

NEURAL PRECURSORS OF ACTION

The control of behavior is a diverse topic that can be examined in myriad different ways. It is perhaps best to start our investigation with the simplest type of behavior control, the decision to make simple actions or to inhibit those same simple actions. Every moment of every day we act in many small ways, and even the most complicated action, such as giving a speech, consists of many smaller actions, such as deciding to emphasize a particular word or to inhibit another word. Some of the most interesting studies of the control of simple behavior are the neuropsychological studies of Benjamin Libet (1981). It was well known at the time that Libet began his research that prior to motor movement there was an electrical change on the area of the scalp above the premotor cortex. This is known as the *readiness potential*.

The readiness potential precedes movement by up to one second. Although this phenomenon had been thoroughly researched, Libet was the first to investigate the relation of the timing of the readiness potential to the conscious thought to make a motor movement. Libet set out to investigate when his subjects became consciously aware of an intention to act and whether the intention to act came before or after the readiness potential. In his experiments, subjects sat still while an electrode was attached to their scalp. They were then instructed to move their hand at random intervals throughout the experiment. He instructed his subjects to note the time when they first became conscious of the intention to move their hand. To measure the time accurately, Libet used a clock that was designed specifically for the experiment, with a hand that moved across the clock face at a speed that allowed subjects to precisely judge the time of occurrence of their intentions. Pilot tests to determine the accuracy of this measurement system were conducted by giving the subjects mild shocks and instructions to estimate the time at which they received those shocks. Surprisingly, subjects were accurate to within 50 ms.

Confident in the accuracy of his time measurement system, Libet began the experiments. Each time a subject made a hand movement, the readiness potential for the seconds prior to that movement was recorded and the subject was asked at what point on the "clock" he or she had first become aware of the intention to move his or her hand. The relative time of occurrence of the readiness potential, the conscious intention to move, and the hand movement could then be calculated. Most people believe that consciousness controls neural activation (thoughts) and movement (action), so they would believe that conscious thought to move the hand should come at the same time as, or slightly before, the readiness potential. This was not the outcome. The readiness potential preceded conscious awareness of intention by about 350 ms, on average (after correcting for error in determining awareness). The awareness of intention preceded action by 150 ms.

What does this tell us about free will? First, these data contradict the naïve view of free will—that conscious intention causes action. Clearly conscious intention cannot cause an action if a neural event that precedes and correlates with the action comes before conscious intention. This does not doom the concept of free will, however. Libet himself argued that a form of free will was supported by this data (Libet, 1999). Although conscious intention did not precede nor was it coincident with the onset of the readiness potential, it still preceded the hand motion by 150 ms. That leaves plenty of time for a person to inhibit the hand motion after they have become aware of the intention to move the hand. This indicated to Libet that while we may not have a free will, we do have a "free won't"—we can stop behavior that has been initiated. Libet's conception of inhibition of a response as the locus of free will leads us to further examine research into inhibition.

Libet's research and ideas have been revolutionary in the fields of both philosophy and neuroscience. He has likewise attracted numerous critics. Below, we briefly detail recent relevant criticisms and related research. We note that far more criticism (Gomes, 2002; Klein, 2002; Oakley & Haggard, 2006; Pockett, 2004, 2006; Pollen, 2004) has been heaped upon Libet's work with the timing of perception (e.g., Libet et al., 1964) than upon his work on readiness potentials described above (for replies to the above criticisms, see Libet, 2002; Libet, 2006). Even one of Libet's most vocal critics considers his finding that readiness potentials precede conscious awareness of intention to be unassailable (Pockett, 2004).

We find the arguments of certain other researchers and philosophers, although less critical of Libet, to be more damaging to his finding. One group of researchers conducted experiments similar to those of Libet, but with subjects who were hypnotized to believe that they were not controlling their actions (Haggard, Cartledge, Dafydd, & Oakley, 2004). This group was compared to a group of subjects who knowingly controlled their actions and to a group of subjects who did not control their actions. The action being measured in all cases was moving a finger to depress a button. The subjects who controlled the button presses were instructed to press the button at random times, whereas the button depressed itself randomly for the subjects who did not control the button presses. Those who were hypnotized were told that the button would depress itself randomly, but they actually chose when to press it. All subjects rated each button press on a scale of how voluntary or involuntary the action was. They also estimated the time when their finger moved, allowing the experimenters to calculate error in perceived time of movement relative to the actual time of movement. Subjects who were hypnotized perceived their finger movements to be involuntary, just like those subjects whose finger movements were involuntary. Also, nonhypnotized subjects who made voluntary movements showed much greater anticipation of their movements, as shown by an earlier perceived time of action relative to the actual time of action, than did subjects who made involuntary movements. However, hypnotized subjects making voluntary movements showed the same amount of anticipation as those who made involuntary movements. This outcome indicates that the hypnotized subjects were not conscious of their free actions.

This research raises interesting questions that we cannot (for reasons of space) address here: If an action can be free when one believes it to be forced, is conscious volition irrelevant to the question of free will? This issue is related to Daniel Wegner's discovery that the feeling of self-control of behavior can be dissociated from actual causes of behavior. That is, people may cause an action but not be aware of it, or they may think they caused an action that was actually triggered by external forces (Wegner, 2003; also see his chapter in this book).

Whereas Wegner uses these findings to argue against the concept of free will, others have argued that even actions that stem from unconscious volition can represent evidence of free will, provided that the actions were not externally caused (Rosenthal, 2002).

Accepting that free will can be unconscious also defuses another criticism of Libet's research: the idea that the "free won't" would not be evidence of free will if preceded by an unconscious neural pattern similar to the readiness potential. If we accept that free will can be unconscious—although Libet (2006) himself refuses to do so—the possibility of an unconscious cause to inhibit action would not cause us to discard the concept of free will. These are thorny issues. We will leave to others the question of whether a conscious decision is necessary for free will or whether an unconscious decision can constitute free will. We do note, however, that the notion of unconscious volition is quite far removed from a straightforward conception of free will.

INHIBITION OF SIMPLE ACTIONS

The simplest method of investigating inhibition is the stop-signal paradigm, pioneered by Gordon Logan about 3 decades ago. In a standard stop-signal experiment, subjects perform repeated trials of a simple task (usually discriminating X from O). On a portion of the trials (usually around 20%), a tone is emitted at some point after the go stimulus (the X or O) has been presented but before the subject has responded to that stimulus (Logan, 1994). Subjects are instructed to stop performing the discrimination task (the go task) when they hear the tone (the stop signal). The go reaction time is the time from when the stimulus has been presented on a go trial until the subject makes a response. So in discriminating X from O, the time that the letter first appears on screen constitutes the beginning of the go task, and the time that elapses between the presentation of the letter and when the subject presses a key to respond constitutes the go-signal reaction time. The stop-signal reaction time is the time from when the stop signal was presented until the stopping process has finished. This time cannot be directly measured, because when the stopping process has ended nothing happens. The length of the stopping process can be estimated, however, by finding the area under the go reaction time distribution to the left of the probability of stopping (Logan, 1994). For example, if the probability of stopping on a stop-signal trial is 85%, and 85% of go-signal reaction times are less than 300 ms, then the time between the go signal and the stop signal is 300 ms. Subtracting the stop-signal delay gives the estimated stop-signal reaction time (Logan, 1994).

Researchers have investigated the effects of different tasks and stop-signal delays on stop-signal reaction time; they have also investigated subject population differences in the task. As we are using Libet's work on the inhibition of prepared responses (as determined by the presence of readiness potentials) as a justification for examining inhibition, we should point out that there has been some research looking at readiness potentials in the stop-signal paradigm (De Jong, Coles, Logan, & Gratton, 1990). This research found that, within the stop-signal paradigm, actions could be stopped even after a readiness potential had begun. This indicates that research on the stop-signal paradigm can inform the question of whether humans have "free won't" and what the limits of that ability are.

Before describing other experiments, we should briefly explain a framework for looking at the data. In this framework, known as the *horse-race model*, there are two competing mental processes: the *go process* and the *stop process*. If the stop process is completed sooner than the go process, the action will be inhibited. Otherwise, the action will be performed. One key assumption of this model is that the two processes are independent. The data generally conform to this assumption (Logan, 1994). Following from this model, there are only four factors that determine whether a response will be stopped: the delay between the beginning of the go task stimulus and the stop signal (the stop-signal delay), the mean reaction time to complete the stop process, the mean reaction time to complete the go task, and the variance of the reaction time to complete the go task (the stop-process variance would also matter, but it is generally assumed to be zero for simplicity). To restate these factors, increased delay between the go signal and the stop signal, faster responses to the go task, and slower responses to the stop signal will lead to a decreased probability of stopping (Logan, 1994). Increased variability of the speed to complete the go task can either increase or decrease the probability of stopping. All experimental variables that influence the probability of stopping do so by affecting one or more of the above factors. We emphasize that the relative finishing times of the go and stop processes—not the relative starting times—determine whether or not an action will be stopped.

One of the first variables for which its relationship to stop-signal reaction time was tested was the delay between the presentation of the go stimulus and the stop signal. Data from an early experiment that tested this relationship are shown in the accompanying figure. These data show the measured reaction time to complete a go trial (left side of the figure) and the estimated stop-signal reaction time as a function of stop-signal delay (right side of the figure). Stop-signal reaction time declines as the stop-signal delay increases, and the go reaction time is longer than the stop-signal reaction time. That is, subjects are faster to respond to a stop signal than to complete the go task, and they respond more quickly to a stop signal as the delay between presentation of the go stimulus and the stop signal increases. This finding has been replicated (Logan, Cowan, & Davis, 1984).

Another consistent finding of stop-signal experiments is that subjects take about the same amount of time to inhibit a wide variety of different actions

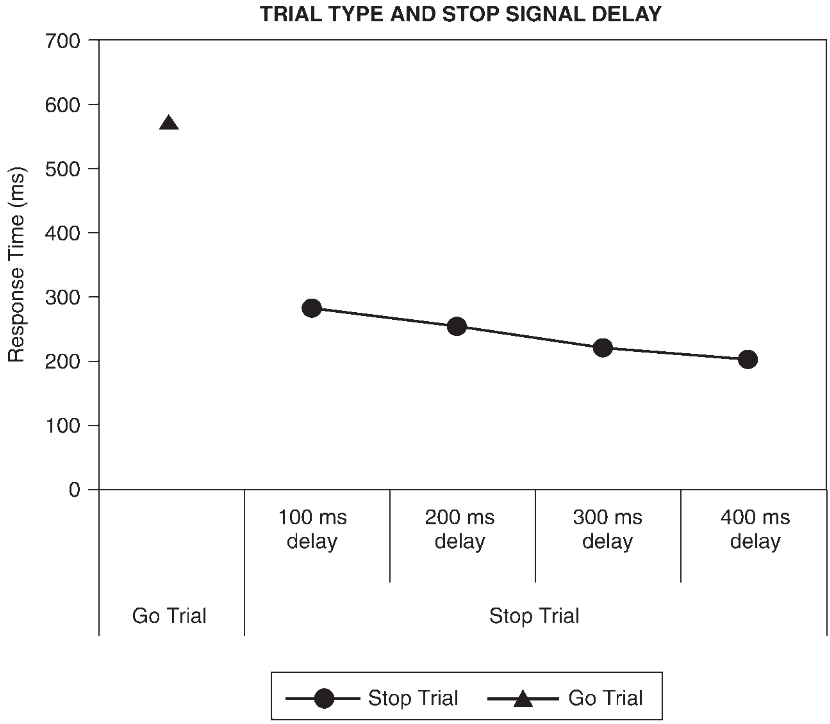

Figure 10.1.

(Logan & Cowan, 1984). People can inhibit both discrete tasks (such as discrete trials of discriminating among letters) and continuous tasks (such as typing or arm waving) with equal ease. Intersubject variability in stop-signal reaction times is also rather small (Logan & Cowan, 1984). Typically, most subjects can stop an action in about 200 ms.

Besides investigating the effects of various task parameters on stop-signal reaction times, researchers have also investigated the differences in stop-signal reaction times in different subject populations. There are numerous patient populations (e.g., those with frontal damage) that have a reduced ability to inhibit in daily living, so it is interesting to know if they show reduced ability to inhibit even in the simple stop-signal paradigm. This type of experiment has been performed with college students who score high in impulsivity (Logan, Schachar, & Tannock, 1997), children with attention deficit/hyperactivity disorder (ADHD) both on and off stimulant medication (Bedard et al., 2003), patients with frontal lobe damage (Dimitrov et al., 2003), and with healthy older adults (Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Rush, Barch, & Braver, 2006).

College undergraduates who scored high on a measure of impulsivity showed a lengthened stop process relative to undergraduates who scored in the normal range on impulsivity (Logan et al., 1997). Because impulsivity did not affect the response time to complete the go task, the effect is not due to a general slowing of cognitive processes. Children with ADHD had longer response times than normal children on stop trials, although they also had more errors on go trials. Those who were medicated with stimulant medication to treat the ADHD reacted more quickly and with fewer errors on both go trials and stop trials (Bedard et al., 2003). These data indicate that failure of inhibition is only one of the problems in children with ADHD. Patients with frontal lobe damage (specifically, frontal lobe lesions and frontal lobe dementia), although slower overall at performing stop and go tasks than control subjects, showed no differential impairment of inhibition in the stop-signal task (Dimitrov et al., 2003). This outcome seems surprising, given that frontal patients show reduced inhibition across many other tasks. Healthy older adults consistently exhibited slower stop-signal reaction times than younger adults (Rush et al., 2006). Although older adults were slower than younger adults on go trials, their slowing relative to younger adults on stop-signal trials was significantly greater than their slowing on go trials (Kramer et al., 1994), revealing a deficit in inhibitory control.

What can we conclude about the control of simple tasks through inhibition? Certainly, we can say that inhibition is a robust human ability. Humans can inhibit all manner of thoughts and actions (whether discrete or continuous). The time course of inhibition is also remarkably similar across different tasks and paradigms. Simply put, the ability to inhibit is one of the most important abilities that humans possess. Without strong inhibitory abilities, humans would be incapable of many of the most rudimentary forms of control over their own behavior (e.g., Hasher, Stoltzfus, Zacks, & Rypma, 1991).

ASSESSING CONTROL IN MEMORY PERFORMANCE

Although the above-mentioned findings highlight the role of inhibition in the control of human behavior, we must still keep in mind that these experimental paradigms take into account performance in only relatively simple tasks of very short duration. They do not directly speak to a much wider range of complex, deliberate actions that are typically longer in duration. Examples might include deciding what to say or not to say in a conversation, what answers to write in a classroom exam, or what information to recall as a witness in the courtroom.

Notice that in all of these types of situations, one can reasonably assume that individuals typically do not simply say or write everything that is on their minds. Rather, they make, or at least have the potential to make, responses in a conscious, deliberate manner, choosing what information to volunteer or withhold, depending on their circumstances and personal goals. The assumption that we have the capacity to carefully decide what and how much to say is especially critical when taking an oath to "tell the truth, the whole truth, and nothing but the truth" in a court of law. Furthermore, by presupposing that individuals have control in deciding how to respond in these types of situations, one might further assume that whatever information comes to mind can vary in terms of the subjective experience associated with that information.

When trying to remember a past event, a person might be able to consciously recollect specific details regarding its context, or specific emotions that were experienced at the time, and be confident that this information stems from the occurrence of that specific event. Or perhaps certain details may unexpectedly come to mind that seem very familiar, but the person cannot consciously recollect the source of this information—and yet, he or she may still choose to recall or not to recall such details as part of the past event due to his or her sense of familiarity alone. In light of these possibilities, it is tempting to think that we have a great deal of control over our thoughts and actions in daily life, especially when it comes to remembering the past. But how much control do we really have under these types of circumstances to remember and communicate information accurately and completely?

At first glance, a cognitive psychologist would be the last person one would turn to for an answer to this question. After all, in laboratory settings, subjects are typically not allowed much leeway in their behavior. What types of responses a subject makes in an experiment are defined and carefully controlled by the experimenter. Responses that fall outside of the acceptable range are corrected in some fashion or considered outliers and subsequently ignored. However, in recent years we have seen a surge in interest in how to empirically study the influence of control processes in a wide range of cognitive tasks. Two influential paradigms concerned with the study of control in human memory performance that we will consider are the process-dissociation procedure, developed by Larry Jacoby and his colleagues, and the free and forced report procedure developed by Asher Koriat and Morris Goldsmith. Both experimental paradigms may be useful to the study of free will in that they provide quantitative estimates of the influence of cognitive control in a range of memory tasks. More importantly, research using these paradigms has helped to identify a number of variables that influence cognitive control.

The process-dissociation procedure (PDP) has been used to estimate the separate contributions of consciously controlled and nonconsciously controlled, or automatic, processes to performance on memory tasks (for a review, see Kelley & Jacoby, 2000). The process-dissociation procedure (Jacoby, 1991) is based on two theoretical assumptions. The first assumption is that performance on a cognitive task does not reflect the operation of a single mental process, but is rather the product of multiple processes operating conjointly. Second, the contributions of the processes to memory performance are independent, which is to say that they can be dissociated under certain experimental conditions or among subject populations. Indeed, the PDP approach directly challenges the notion that human beings have complete control over their behavior. For the types of thoughts and actions that are typically analyzed using PDP, such as performance on memory-related tasks of recall, recognition, or completing word stems (e.g., ele______) under various instructional sets, one can, at best, only exert partial conscious control.

To actually implement PDP, one must use an *opposition procedure* (Jacoby, Woloshyn, & Kelley, 1989) in which the two types of mental processes (e.g., conscious vs. unconscious, controlled vs. automatic) that are presumably tapped by a given cognitive task are set in opposition to each other. For example, in one type of experiment, subjects might first study a list of words (e.g., *element*) and then perform a word-stem completion task (e.g., ele_______). In one testing condition, the "inclusion" condition, subjects are asked to complete the word stems using only previously studied words (so ele________ would be completed by *element* if the subject is successful). In the second "exclusion" condition, subjects are told to complete stems with words that were *not* previously studied (so *elephant,elegant,election,* or *electric,* etc., would be appropriate). When subjects mistakenly complete a stem with a studied word in this exclusion condition—they produce *element,* contrary to instructions—their performance suggests the influence of automatic memory processes. That is, if production of *element* is elevated relative to a baseline condition in which the word had not been studied, the production must occur because the word was primed or activated by prior study but could not be successfully opposed or inhibited by conscious recollection of the event (a controlled process). The PDP, by assuming the independence of controlled and automatic memory processes, is able to provide quantitative estimates of the separate contributions of the two process types based upon performance in both inclusion and exclusion conditions. Details of the computations are outside the scope of this chapter, but can be found in Jacoby, Toth, and Yonelinas (1993), among other places.

The PDP has shown, for instance, that divided attention at study significantly reduces conscious recollection, but leaves unconscious or automatic influences on memory unaffected (Jacoby et al., 1993). Along similar lines, Hay and Jacoby (1996) used a variant of PDP to study the effects of experimentally trained habitual behavior and conscious recollection on cued recall performance. In an initial training phase, subjects studied words that were paired with typical responses 75% of the time (e.g., knee-bend) and atypical responses 25% of the time (e.g., knee-bone). Next, subjects studied a list of word pairs, some of which included words paired with atypical associates (knee-bone, in this context). During a final test, subjects were shown one member of each word pair and a fragment of its associate (e.g., knee-b_n_) with instructions to recall the dominant response from the first list.

Under these conditions, the authors assumed that correct recall of typical pairs (knee-bend) could be based on either conscious recollection for the list or on subjects' trained habitual response (an automatic influence). Conversely, when subjects mistakenly recalled the habitual response after having studied an atypical pair, the authors assumed that habit was the basis for the response. By further assuming that these two sources for responding were independent, Hay and Jacoby (1996) calculated separate estimates of the contributions of conscious recollection and automatic habit on cued-recall performance. They found that whereas varying the amount of initial training of habitual responses did not affect the level of contribution of conscious memory in cued recall, it did affect the estimates of habit. In fact, estimates of habit (automatic responding) directly corresponded to the amount of initial habit training. Conversely, varying list presentation rate during study and cued recall response time during the test affected estimates of consciously controlled recollection, but did not affect estimates of habit. More recently, Hay and Jacoby (1999) administered this procedure to college students and older adults and demonstrated that cued-recall performance differed between age groups only because older adults' ability to consciously recollect the study list was impaired. However, the contribution of habitual responding was age invariant.

Taken together, these and a host of related studies have provided support for the notion that different memory processes can make independent contributions to performance on a single task. More importantly, these results appear to affirmatively answer the main question of free will proposed in this chapter, which is whether human beings exercise control over their actions and decisions. How much control human beings can exercise on any given task is another question altogether. To the extent that free will is related the control of human behavior, the PDP offers a unique approach to the study of free will. By assuming that performance on any given cognitive task cannot be the result of a single, controlled thought process, the PDP approach suggests that whatever behaviors one might consider pure acts of free will are really the product of controlled and automatic process or conscious and nonconscious thoughts and actions. Most important behaviors as analyzed by the PDP reveal the behaviors to be partly automatic and partly under conscious control. If one embraces a notion of free will that can accommodate such "process impurity" in determining human behavior, then the PDP offers a powerful analytic tool for measuring the relative contribution of controlled or conscious processing as well as identifying the factors that influence our control.

MANIPULATING OPTIONS TO REPORT MEMORIES

Another highly influential experimental approach to the study of control of human behavior permits subjects a larger degree of personal control to regulate both the quantity and quality of their responses in cognitive tasks. In typical memory experiments, subjects are told how to respond (recall studied material in order or recall it in any order) and they are often told whether or not to guess. A newer line of research, pioneered by Koriat and Goldsmith (1994, 1996), focuses on the role that *report option*, the decision to volunteer or withhold information, plays in determining performance on memory tests. The basic idea is to test people under various conditions in which they are encouraged or forced to guess with other procedures in which they can withhold answers if they are unsure about the response. In an initial study, Koriat and Goldsmith (1994) compared the use of free and forced report procedures on both the quantity of memories produced and their accuracy, and in both recall and recognition tasks. Subjects attempted to answer general knowledge questions in either a recall format (e.g., What was the name of the composer who wrote the *Moonlight Sonata?*) or a recognition format (e.g., In response to the last question, choose among the following alternatives: Beethoven, Bach, Tchaikovsky, Schumann, Brahms). In addition, subjects were given either free report (i.e., respond to whichever items you believe you can answer) or forced report (i.e., respond to each and every question) instructions. Such a procedure seems particularly relevant to the study of free will, and comparison of performance under the two instructional sets should prove informative. Free report gives subjects the freedom to choose which questions to answer and which information to provide as answers, whereas forced report performance does not permit such freedom but requires subjects to respond (so both conscious and automatic influences may be operating).

The results showed that the use of forced report had no effect on increasing the quantity of correct responses reported relative to free report. This is somewhat surprising, because one might have expected subjects to come up with additional correct answers in the forced report condition, either by producing answers that were below a threshold for recall or just by random guessing alone. By contrast, the use of free report enhanced memory accuracy (the proportion of total responses made that were correct) in both recall and recognition (Koriat & Goldsmith, 1994, Experiment 1). This pattern of results also extended to performance on recall and recognition of studied word lists, a more conventional set of laboratory-based tasks (Koriat & Goldsmith, 1994, Experiment 2).

What these and earlier studies highlight concerning control of responding is that people appear to have little control over the amount of information they can retrieve at any given time. Changing recall criteria by encouraging subjects to make more responses does not lead to increases in amount of information a person can correctly remember, relative to standard free recall instructions with a warning not to guess (Bousfield & Rosner, 1970; Roediger & Payne, 1985; Roediger, Srinivas, & Waddil, 1989). In other words, allowing subjects more or less control in their memory reporting does not affect *how much* accurate information they can recall. By contrast, people *can* exercise some control over the accuracy of the information they do choose to recall when they are permitted to pass or to omit erroneous answers.

More important, a third experiment by Koriat and Goldsmith (1994) tested whether varying motivation for response accuracy would lead to further improvements in accuracy. While attempting to answer general knowledge questions with free report instructions, subjects were given either moderate or strong incentives to provide only correct responses. Under moderate incentive conditions, the reward for correct responding was equal to the penalty for incorrect responding. Under high incentive conditions, subjects risked being penalized several times more heavily for committing errors than for being correct. The authors found that accuracy could be improved with increased incentives, but that the boost in memory accuracy came at a cost to memory quantity, as fewer correct responses were made. Koriat and Goldsmith (1994) argued that this quantity-accuracy trade-off occurred because accuracy can be improved by withholding answers, but because the process of covertly screening potential responses, or monitoring effectiveness, is not perfect, a cost is incurred in terms of decreased quantity of responses.

In subsequent work, Koriat and Goldsmith (1996) examined how the additional factors of subjective confidence and monitoring effectiveness influence control of memory performance, and they proposed a theoretical framework to account for their collective findings. Whereas subjective confidence involves monitoring the accuracy of potential responses as they come to mind, monitoring effectiveness refers to one's ability to discriminate between correct and incorrect answers. Subjects first attempted to answer general knowledge questions under forced report conditions in both tests of recall and recognition, and then they assigned confidence ratings to their answers. In the next phase, they attempted to answer the questions under free report conditions with either moderate or high accuracy incentive.

Just as in their previous work, Koriat and Goldsmith (1996, Experiment 1) observed a quantity-accuracy trade-off in that subjects could achieve a higher level of accuracy in their free recall and recognition of answers to general knowledge questions when given a stronger motivation to be accurate, but at the cost of withholding a greater number of correct answers. In addition, subjects tended to volunteer responses in accordance with their confidence in the accuracy of those responses. Koriat and Goldsmith interpreted this finding to suggest that when remembering past events under conditions of free report, individuals apply a control threshold that allows for the output of responses that have the highest subjective probability of being correct. If a response does not surpass this threshold, then it will be withheld. And, as already mentioned, memory accuracy can be improved by increasing the motivation for accuracy, because such an increase encourages a person to set a higher control threshold.

Of course, the idea that judgments are based on a response criterion is not new. For instance, signal detection theory (SDT; Green & Swets, 1966; Wixted & Stretch, 2004), which is often utilized to analyze recognition performance data, assumes that an individual sets a response criterion, calling items whose familiarity or strength exceeds a response threshold "old" and items whose familiarity falls below the response criterion "new." However, the utility of SDT is generally restricted to forced-report testing conditions in which a subject must respond either "old" or "new" to each and every test item. By contrast, the forced and free report procedure allows for the measurement of changes in response criteria that occur under both forced and free report conditions. Tulving (1983) also proposed that people have "conversion thresholds" that operate to determine whether they will report a retrieved event as a memory, with some conditions of responding leading to more stringent thresholds than others.

With regard to monitoring effectiveness, one can imagine that having a poor ability to distinguish between right and wrong answers would undermine a person's ability to improve accuracy in memory reporting. Indeed, when subjects attempted to answer "deceptive" general knowledge questions such as, "What is the capital of Australia?" (hint: the correct answer is *not* Sydney), monitoring effectiveness proved to be quite poor, and as a result, increased control of memory reporting yielded little or no benefit (Koriat & Goldsmith, 1996, Experiment 2). In other words, subjects continued to respond on the basis of their confidence in the accuracy of retrieved information, except that due to the deceptive nature of the questions, subjective confidence was no longer a reliable guide for accurate responding. For questions that were not deceptive, monitoring effectiveness improved, and as a consequence, greater increases in accuracy were observed at lower costs in memory quantity. In this case, subjective confidence served as a more reliable, albeit imperfect, response index (Koriat & Goldsmith, 1996, Experiment 2). The authors argue that, in theory, when monitoring effectiveness is perfect, there should be no quantity-accuracy trade-off whatsoever. (By the way, Canberra is the capital of Australia).

By contrast, SDT treats subjective confidence as being synonymous with memory strength in that confidence judgments are used to construct ROC (receiver operating characteristic) curves by assuming that confidence directly corresponds to memory strength. Thus, one advantage of Koriat and Goldsmith's theoretical framework is that it can take into account situations in which subjective experience is not a reliable basis for remembering. Consider the use of eyewitness testimony to identify suspected criminals from lineups. Eyewitnesses who are very confident in their identification can provide very compelling evidence in a court of law. And yet, eyewitness research has shown that subjective confidence may not be a reliable guide for accurate identification (for a brief review, see Wells, Olson, & Charman, 2002). For instance, providing reinforcement to eyewitnesses during in a lineup procedure (e.g., "Good job. You are a good witness") may increase their confidence in their response without improving its accuracy (Wells & Bradfield, 1999). Also, repeated questioning of eyewitnesses about mistaken memories does not lead them to revise their recollections. Instead, it leads eyewitnesses to inflate the confidence of their false recollections (Shaw, 1996; Shaw & McLure, 1996).

More recently, Goldsmith and Koriat (1999) have examined an additional means of subject control in memory reporting—control over "grain size," which is the level of generality or detail of a response. For instance, in response to the question, "At what time did the robbery take place?"—rather than simply volunteering or withholding a response, an individual might choose to frame the answer in a way that is more likely to be accurate (e.g., "in the late morning" as opposed to "at 11:30 A.M."). At the same time, relying too much on accuracy may render the reported information uninformative, such as when a person is asked when World War II occurred, and, as a response offers, "Sometime in the 20th century." Indeed, research shows that the choice of grain size is influenced by a person's attempt to compromise between competing tendencies to be both accurate as well as informative in a given situation (Goldsmith & Koriat, 1999; Yaniv & Foster, 1995).

According to Koriat and Goldsmith, the control that individuals exhibit in their memory reporting may be represented by a control threshold for responding, the setting for which will vary depending upon a person's goals and circumstances. To quantitatively assess the factors that influence this control process, Koriat and Goldsmith (1996) developed an analytic procedure, the quantity-accuracy profile, or QAP. QAP describes the joint levels of quantity and accuracy performance that might be attained at various control thresholds, depending upon an individual's overall level of retention and monitoring effectiveness. In addition to providing standard quantity and accuracy measures or memory performance, the QAP provides estimates of monitoring effectiveness and control, and encourages researchers to take into consideration subjects' goals and incentives in experimental situations. To date, though, the method does not accommodate control over grain size in memory reporting.

Koriat and Goldsmith's (1996) framework has been applied to subject populations, such as children, older adults, and schizophrenia patients, who exhibit deficits in memory performance to determine the extent to which such deficits may be attributed to impaired control processes (Danion et al., 2001; Kelley & Sahakyan, 2003; Jacoby et al., 2005; Koriat et al., 2001; Meade & Roediger, 2006; Rhodes & Kelley, 2005). For instance, it has been shown that children as young as 8 to 9 years old can take advantage of free report to increase the accuracy of their memory reports. However, such children could not achieve the level of performance, both in terms of quantity and accuracy, exhibited by children several years older (Koriat et al., 2001). Similarly, older adults show smaller gains in accuracy performance relative to younger adults when they shifted from forced to free report testing conditions (Jacoby, Bishara, Hessels, & Toth, 2005; Kelley & Sahakyan, 2003; Meade & Roediger, 2006; Rhodes & Kelley, 2005).

For our purposes, it is worth noting that the form of control that individuals appear to exercise in the free and forced report procedure is the capacity to suppress or inhibit already retrieved candidate responses at a relatively late processing stage, just prior to verbal report. Candidate responses may be retrieved through conscious effort, nonconscious or automatic processes, or both. The theoretical framework of Koriat and Goldsmith does *not* assume that information can only be initially retrieved in a nonconscious or automatic manner. Indeed, Jacoby and colleagues have recently shown that people can control what information comes to mind during retrieval, a process they term *early selection* (e.g., Jacoby, Shimizu, Daniels, & Rhodes, 2005). By contrast, the role of the control mechanism in the Koriat and Goldsmith framework is to serve as a gatekeeper, after responses have been generated or selected. Above-threshold retrieved information is passively allowed to reach the stage of verbal report; sub-threshold retrieved information is suppressed or inhibited from output. This framework challenges the notion that human beings have the ability to exercise full conscious control of their behavior on memory-related task performance. Rather, Koriat and Goldsmith (1996) propose a more modest domain for the control of human memory.

CONCLUSIONS

We have reviewed four bodies of evidence that deal with the issue of conscious control of behavior. We have danced around the issue of whether conscious control is to be equated with free will; in fact, we suspect that at the most basic level, the answer must be no. Even behavior that subjects believe to be completely under conscious control is influenced by external factors, as discussed above. Yet even if we cannot draw firm conclusions about free will from the body of research that we have summarized, we do think that the four areas of research we have reviewed are relevant to the issues at hand. However, the work in each area of research may raise as many questions as it settles at this point. For example, the ability to inhibit responses is powerful and one could reasonably make the argument that without it, free will would not be possible, because we would not be able stop what some force (external or internal) seemed to impel us to do. Unfortunately, this still leaves unanswered the question of whether we have free will, despite our demonstrated abilities to consciously inhibit behavior. If we do assume that free will can be directly observed in inhibition of behavior, then is free will measured by inhibition?

The latter question is especially poignant with respect to those people who show marked inhibition deficits, such as children, the elderly, and certain patient populations. If we agree with Libet (1999) and consider inhibition of an unconsciously determined action to be equivalent to free will, then these populations suffer from impaired free will. At first glance, it seems quite odd to view free will as a measurable mental ability and to characterize broad swaths of the population as having impaired free will. Upon further consideration, though, we see that this view is already implicitly, if not explicitly, accepted in many fields, as well as (to some degree) enshrined in common wisdom about behavior. This principle is most visible in how the legal system treats offenders depending upon whether or not an act was premeditated. In the United States and in most other countries, legal systems consider premeditated murder to be a much worse offense than murder that is not premeditated, even if the circumstances of the murder are otherwise identical. This accepted principle shows that society (or the legal system) considers people to have less free will when they are in a fit of rage than when they plan carefully and laboriously to commit an act.

Returning to the work of Koriat, Goldsmith, and colleagues, we can also ask similarly difficult questions. For instance, should children, older adults, or for that matter, any one of us, serve as witnesses and take oaths to tell "the truth, the whole truth, and nothing but the truth" if we fully recognize the difficulties of monitoring memory accuracy and the severe consequences for reporting false memories? Again, if the locus of free will is in the control of behavior, then those who can exercise less control of certain behaviors have, by this logic, less free will. If people show improved control in a task with practice, are we justified in saying that they have enhanced free will? Although this idea sounds bizarre when stated plainly, voluntary control as directly related to free will is a common concept. Belief in the ability of "self-control," whether in resisting eating too much or driving too fast or drinking alcohol, is widespread. One reason people consult clinical psychologists is to help them gain self-control (in weight management, in alcohol consumption, or in managing their level of anxiety or depression). We have focused on approaches from experimental psychology in this chapter, but certainly issues in the clinical settings arising from the self-control exerted (or not exerted) by patients is another relevant arena in considering free will. The issue of free will must always be part of psychology, even if it is never completely settled.

REFERENCES

Bedard, A.-C., Ickowicz, A., Logan, G. D., Hogg-Johnson, S., Schachar, R. J., & Tannock, R. (2003). Selective inhibition in children with attention-deficit hyperactivity disorder off and on stimulant medication. *Journal of Abnormal Child Psychology, 31*, 315–27.

Bousfield, W. A., & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science, 20*, 75–76.

Danion, J. M., Gokalsing, E., Robert, P., Massin-Krauss, M., & Bacon, E. (2001). Defective relationship between subjective experience and behavior in schizophrenia. *American Journal of Psychiatry, 158*, 2064–66.

De Jong, R., Coles, M. G. H., Logan, G. D., & Gratton, G. (1990). Searching for the point of no return: The control of response processes in speeded choice reaction performance. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 164–82.

Dimitrov, M., Nakic, M., Elpern-Waxman, J., Granetz, J., O'Grady, J., Phipps, M., et al. (2003). Inhibitory attentional control in patients with frontal lobe damage. *Brain and Cognition, 52*, 258–70.

Free will. (2006). Retrieved October 3, 2006, from http://en.wikipedia.org/wiki/Free_will

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. In D. Gopher & Koriat (Eds.), *Attention and performance XVII—Cognitive regulation of performance: Interaction of theory and application* (pp. 373–400). Cambridge, MA: MIT Press.

Gomes, G. (2002). Problems in the timing of conscious experience. *Consciousness and Cognition, 11*, 191–97.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Haggard, P., Cartledge, P., Dafydd, M., & Oakley, D. A. (2004). Anomalous control: When 'free-will' is not conscious. *Consciousness and Cognition, 13*, 646–54.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and Inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 163–69.

Hay, J. F., & Jacoby, L. L. (1996). Separating habit and recollection: Memory slips, process dissociations, and probability matching. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 1323–35.

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and elderly adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging, 14*, 122–34.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–41.

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General, 134*, 131–48.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*, 852–57.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139–54.

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*, 115–25.

Kelley, C. M., & Jacoby, L. L. (2000). Recollection and familiarity: Process-dissociation. In E. Tulving and F.I.M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 215–28). New York: Oxford University Press.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48*, 704–21.

Klein, S. A. (2002). Libet's temporal anomalies: A reassessment of the data. *Consciousness and Cognition, 11*, 198–214.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297–315.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology, 79*, 405–37.

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and Inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9*, 491–512.

Libet, B. (1981). The experimental evidence for subjective referral of a sensory experience backwards in time: Reply to P.S. Churchland. *Philosophy of Science, 48*, 182–97.

Libet, B. (1999). Do we have free will. *Journal of Consciousness Studies, 6*, 47–57.

Libet, B. (2002). The timing of mental events: Libet's experimental findings and their implications. *Consciousness and Cognition, 11*, 291–99.

Libet, B. (2006). The timing of brain events: Reply to the "special section" in this journal of September 2004, edited by Susan Pockett. *Consciousness and Cognition, 15*, 540–47.

Libet, B., Alberts, W. W., Wright, E. W. J., Delattre, L. D., Levin, G., & Feinstein, B. (1964). Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *Journal of Neurophysiology, 27*, 546–78.

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In E. D. Dagenbach & T. H. Carr (Ed.), *Inhibitory Processes in Attention, Memory, and Language.* San Diego: Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Science, 91*, 295–327.

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 276–91.

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science, 8*, 60–64.

Meade, M. L., & Roediger, H. L., III. (2006). The effect of forced recall on illusory recollection in younger and older adults. *American Journal of Psychology, 119*, 433–62.

Oakley, D. A., & Haggard, P. (2006). The timing of brain events: Authors' response to Libet's "reply." *Consciousness and Cognition, 15*, 548–50.

Pockett, S. (2004). Hypnosis and the death of "subjective backwards referral." *Consciousness and Cognition, 13*, 621–25.

Pockett, S. (2006). The great subjective back-referral debate: Do neural responses increase during a train of stimuli? *Consciousness and Cognition, 15*, 551–59.

Pollen, D. A. (2004). Brain stimulation and conscious experience. *Consciousness and Cognition, 13*, 626–45.

Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis. *Journal of Memory and Language, 52*, 578–94.

Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory and Cognition, 13*, 1–7.

Roediger, H. L., III, Srinivas, K., & Waddil, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *Journal of Experimental Psychology: General, 118*, 253–57.

Rosenthal, D. M. (2002). The timing of conscious states. *Consciousness and Cognition, 11*, 215–20.

Rush, B. K., Barch, D. M., & Braver, T. S. (2006). Accounting for cognitive aging: Context processing, inhibition or processing speed. *Aging, Neuropsychology, and Cognition, 13*, 588–610.

Shaw, J. S., III. (1996). Increases in eyewitness confidence resulting from postevent questioning. Journal of Experimental Psychology: Applied, 2, 126–46.

Shaw, J. S., III, & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior, 20*, 629–54.

Tulving, E. (1983). Elements of episodic memory. Oxford, U.K.: Oxford University Press.

Wegner, D. M. (2003). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science, 10*, 138–44.

Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science, 11*, 151–54.

Wilson, E. O. (1998). *Consilience.* New York: Alfred A. Knopf.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of Remember/Know judgments. *Psychonomic Bulletin & Review, 11*, 616–41.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124*, 424–32.

11 Self Is Magic

Daniel M. Wegner

Imagine a magician who can make things happen merely by *thinking* of them. This magician thinks "I'd like the lights on," and before you know it…there is light! Right there and then, a hand has reached out and turned on a lamp. Next, the magician hopes for warmth. In a matter of minutes, a fire is glowing and crackling in the fireplace. The magician's wish for a cup of tea and a plate of cookies soon unfolds into just the right tea and just the right cookies, placed conveniently at arm's length. Every wish seems to come true. And when the next cookie comes to mind, the arm reaches to the plate, grasps a cookie (not just any cookie, but the very one that the magician is thinking about) and deftly brings it to the conjurer's mouth at the appropriate angle for a nice bite. What wouldn't we all give to live such an enchanted life?

Aside from the occasional stray cookie that slips to the floor, we do live this life. Human action is a kind of magic, an astonishing ability to think of something and thereby make it happen. Although a lamp will seldom light on its own merely because we want it to, we find that our fingers leap to the switch and light that lamp when the idea comes to mind. Our actions are an astonishing realm of events that bend to our desires when so much of the world does not. Perhaps this is why each person views self with awe—*The Great Selfini* amazes and delights! We are enchanted by the operation of our minds and bodies into believing that we are "uncaused causes," the origins of our own behavior. Each self is magic in its own mind.

Unfortunately, the magic self stands squarely in the way of the scientific understanding of the psychological, neural, and social origins of our behavior and thought. As long as we are charmed by ourselves, perhaps we won't see beyond the magic. This chapter examines this curious standoff in several ways. We will begin by exploring just how it is that the self's magic seems to arise. We will try to "reverse engineer" the magic, discerning what kind of system might be necessary to make us into such impressive beings that we are spellbound by our own performances. Then, we will consider why it is that this concept of self as an inner origin of our actions is so resistant to analysis and understanding—why even those of us who hope to overcome our belief in magic are still captivated every day by the self's parlor tricks. And finally, we will turn to the question of why the process of evolution might have unfolded in such a way as to make us think we are magical creatures.

MAKING MAGIC: THE BIRTH OF AN EGO

How would one go about making a being that believed in its own magic? Could a robot or computer be constructed with this property? This entity would need to have many of the standard cognitive properties of the human. Of course, it would need to be able to perceive events and develop causal theories of their interrelations, and it would need to have the ability to reflect on itself and focus those causal theories on its own processes. These are standard features that are often discussed by those who imagine reverse engineering a human mind (e.g., Angel, 1989; Braitenberg, 1984; Dennett, 1994; Holland, 2003; Scassellati, 2002). But to believe itself magical, this entity would need to have one property we don't often appreciate in the cognitive toolkit of the standard human: *incomplete self-knowledge.* Perceiving magic of any kind requires that we don't fully understand how something has happened.

This realization follows directly from the theory of magic perception introduced (with a flourish) by Harold H. Kelley (1980). According to Kelley's analysis, we perceive magic when an apparent causal sequence shortcuts or obscures a real causal sequence that is not itself fully perceived (and the apparent sequence usually departs from one that common sense would predict). When the magician saws a lady in half, for example, an event has occurred that is only apparent: The lower half of her body seems to have seceded from the upper half. We know this can't be the real causal sequence, or there would be a lot of magician's assistants out there whose careers were tragically cut short. There *is* a real causal sequence that may involve contortion, mirrors, special cavities in the box, an additional lady to provide distal wiggling feet, and what else only the magician knows. But the audience sees the apparent causal sequence in which the saw seems to cut her in half, and if the illusion is well done, they are amazed by the trick. It is in exactly this sense that the self is magic: When we look at ourselves, we perceive a simple and often astonishing apparent causal sequence (I thought of it and it happened!) when the real causal sequence underlying our behavior is complex, multithreaded, and unknown to us as it happens.

Now, saying that people believe in magic, at least in our enlightened times, can be something of an insult. The film *Jesus Is Magic* by comedienne Sarah Silverman, for example, chides Christians about their beliefs and suggests, all in jest of course, that they are being duped somehow. Belief in magic may be fun and even a source of delight, but it also can entail childlike naiveté, delusion, or just plain foolishness. Seeing one's own causal influence as supernatural is part of being human, though, so rather than ruing this human tendency or calling it foolish, it is psychological science's job to understand it. How do people develop this magic self—what Dennett (this volume) calls "some concentrated internal lump of specialness"? Why do we experience our actions as freely willed, arising mysteriously from the self, and why, too, do we resist attempts to explain those actions in terms of real causal sequences, events that are going on behind the curtain of our minds?

One explanation of the magic draws on the idea that the mind presents us with only a relatively impoverished account of its own operations, and our attempt to make sense of the evidence yields the impression that we are freely willing our actions. This account is the basis of the theory of *apparent mental causation*, a set of ideas that draw on the philosophy of Hume (1739/1888) to explain how it is that people come to experience conscious will (Wegner, 2002, 2003, 2004; Wegner & Wheatley, 1999). This theory and several related accounts (e.g., Brown, 1989; Claxton, 1999; Michotte, 1963; Nisbett & Wilson, 1977; Spence, 1996; Thompson, Armstrong, & Thomas, 1998) propose that people experience willing their actions when they draw causal inferences relating their thought to their action. Quite simply, a person infers that an event is due to the self as a result of perceiving a causal link between own thoughts and that event. Hume's insight was to note that such perception is a matter of inference, not direct perception, and the implication of this insight is that the perception of one's own causality is open to error.

How do we go about drawing this causal inference about our own action? Consider that magical act of turning on the light. This is something that sometimes can feel quite willful, and at other times can feel absentmindedly automatic. If you have just thought about turning on the light and then do so, it may feel more willful—whereas if you have been thinking about having a cookie and then suddenly find yourself turning on a light instead, it is likely to feel less willed and more like some sort of alien control. To support a feeling of will, the thought of turning on the light also must occur just prior to the action to maximize the experience of will, as thoughts that occur far beforehand (and that then are forgotten until the action) or thoughts of flipping that switch that only appear after the light is on do not seem to prompt a sense of willed action. And if someone else presses your hand to the lamp, you may discount entirely the causal role of your prior thought and feel the act is unwilled. These observations point to three key sources of the experience of conscious will—the *consistency, priority,* and *exclusivity* of the thought about the action. For the perception of apparent mental causation, the thought should be consistent with the action, occur just before the action, and not be accompanied by other potential causes. Several studies have examined the influence of each of these principles of apparent mental causation.

Consistency Effects

The idea that a thought and action must be consistent with each other to yield an experience of conscious will has been investigated in experiments on magical thinking (Pronin, Wegner, McCarthy, & Rodriguez, 2006). These studies examined whether priming people to experience thoughts consistent with events they did not actually cause might lead them to experience the events as caused by the self.

Participants in one experiment were asked to play the role of a witch doctor in a study of psychosomatic influences on health, and to perform a voodoo curse by sticking pins in a doll in the presence of another participant assigned to play the role of the victim. The victim role was in fact played by an experimental confederate, who later feigned a headache. The question of interest for this research was whether participants would accept any causal responsibility for this headache: Would they believe that they had exerted some influence over the victim's health? A participant would not need to believe in voodoo per se to reach this conclusion, but would only need to perceive that a potential victim might be stressed into a headache by the shock of receiving a curse. Participants were given an article suggesting this possibility in preparation for the experiment (Cannon, 1942), and many did accept the idea that their actions caused the headache.

Magical thinking was amplified in this experiment by a manipulation of the consistency of participants' thought with the action. The inference that "I made the victim sick" was significantly strengthened among participants who were led to dislike the confederate before performing the voodoo curse. These participants were exposed to a confederate who was late, rude, and messy; postexperimental questioning revealed that they had indeed come to dislike this person. As compared to participants who met an unremarkable confederate—one who was normal and likable—those who performed the curse on the confederate victim they disliked were more inclined to believe that their curse had caused the victim's headache. In a follow-up study, the same phenomenon was observed when participants were merely instructed to "think negative thoughts" about the victim before pinning the doll. As compared to those who were not given this instruction, the participants led to think in a way that was consistent with the act of harming the victim came to believe that they had indeed caused such harm. And of course, no harm had been caused at all.

These findings suggest that people can easily develop the belief that they are harming someone when they have wished for such harm, even when the harm befalls the victim for reasons unrelated to the harm-wisher's desires. This may be why we can feel guilty if we have wished ill on someone and they suffer an unfortunate fate—even though we know we really were not responsible. Cursing an obstreperous elderly relative under our breath may lead us to feel particularly culpable when the relative falls and breaks a hip. Our natural tendency to link our thoughts with consistent events leads us all too readily to the mistaken belief that the events have issued from our will.

The self can also be magic in a good way. To study this possibility, we put the voodoo doll in mothballs and turned to the magic of cheerleading. Further studies by Pronin et al. (2006) revealed that there is a tendency to believe that one has caused a positive event to occur merely because one has imagined it. In these experiments, people were led to think about the outcomes of sporting events being played by others, but were questioned afterward about their own causal influence on the sporting outcomes. People who were asked to envision the success of a basketball shooter on each of eight free throws were more inclined to believe they had indeed helped him than those who were asked to envision him lifting barbells when he then proceeded to sink six of the eight shots. People who were watching all this and were informed about what the spectators were visualizing reached the same conclusion: Even these uninvolved onlookers thought the spectator visualizing successful shots had somehow helped the shooter to succeed.

This tendency to claim authorship for the successes of others also extends to one's favorite sports teams in real competition. In another study, fans at a basketball game were asked to complete a pregame exercise—either thinking about how each of their team's players could contribute to the game, or thinking about how each player could be identified in a crowd. When the fans were then quizzed in the middle of the game to see whether they felt they were personally influencing the game's outcome, those who had been prompted to develop success-relevant thoughts for their team were more likely to report exerting influence. A final study by Pronin et al. (2006) found that fans watching the 39th Super Bowl football game on television were susceptible to the same illusion. Those viewers who reported thinking more about the outcome of the game also claimed more personal responsibility for the game's outcome—regardless of whether their favorite team won or lost (the Patriots won, by the way, largely because I had wished this).

These studies highlight what may be a general process of mind underlying belief in paranormal phenomena such as ESP, clairvoyance, precognition, and psychokinesis. In everyday life, our bodies appear to respond readily and easily to many of our wishes. Yes, we may find it difficult to wish to perform a Chopin étude on the piano when we've never taken a lesson—but there are so many things we *can* do, things that happen just because we want them. It makes sense that this normal human capacity for conscious will might lead us into overextensions from time to time. If our wishes seem to prompt a range of activity within our personal sphere of influence, why not hope for more? The many forms of supernatural belief, including beliefs in the effectiveness of appeals to deities, may develop as natural next steps that follow from the magic we perceive in ourselves. If mere wishing can pop the lid off a bottle of beer, why not wish for the moon?

The belief that one is influencing events also can be enhanced by prior action-consistent thoughts that are not conscious. Preaction subliminal primes of action effects can increase perceptions of authorship for the action. This result was observed in a study when people were asked to judge whether their button press was responsible for the resting position of a marker on a computer display (Aarts, Custers, & Wegner, 2004). There were two markers moving very quickly on the display, and the participant's task was to judge whether their marker was the one that had come to that resting spot. On some trials, the resting spot was primed with a brief flash at that position, and it was on these trials that participants estimated more often that the marker was their own—even when the flash priming that position was so brief as to be subliminal. In subsequent research, this effect has also been observed among depressed individuals—suggesting that the consistency of thought and action enhances perceptions of own agency even among people with weakened self-views that surface in depression (Aarts, Wegner, & Dijksterhuis, 2006).

The consistency of thought and action can be undermined even in normal action when people are distracted from their action-consistent thoughts. People in one study were asked to suppress thinking about what they were doing as they performed each of a series of simple activities—such as winding thread on a spool (Wegner & Erskine, 2003). They reported being somewhat successful at the suppression, and also reported weakened feelings of voluntariness for these actions. The active disengagement from thoughts about actions may be a pathway leading to the phenomena of hypnosis, a kind of "voluntary involuntariness" that comes about when people are instructed to ignore their own thoughts about what they are doing (Lynn, Rhue, & Weekes, 1990; Lynn, Weekes, Matyi, & Neufeld, 1988).

Priority Effects

The second principle of apparent mental causation, that self will be seen as causal when the thought of action occurs just prior to the action, has also been tested in research. Priority effects were observed initially by Wegner and Wheatley (1999), and have also been obtained in studies of vicarious agency by Wegner, Sparrow, and Winerman (2004). For these latter experiments, participants were led to experience the arm movements of another person as if the movements were their own. The participant was attired in a robe and positioned in front of a mirror such that the arms of a second person standing behind the participant could be extended through the robe to look as though they were the arms of the participant. The second person wore gloves to aid in this illusion. Participants kept their own arms at their sides and were instructed not to move. Both participant and "hand helper" wore headphones.

For the experiments, the helper's arms performed a series of 32 movements (e.g., snapped the fingers of the right hand, waved hello with both hands) in response to sequential instructions the helper was given via the headphones. In one experiment, participants also heard the instructions for each of the arm movements through their own headphones, or they heard nothing. Those who heard the instructions thus were provided with consistent prior thoughts for actions they perceived visually to be occurring in the position their own actions might occur. As might be expected from the aforementioned consistency studies, the consistent previews led participants to report enhanced feeling of control over the arm movements as compared with other participants. Participants did not feel that they had full control of the arms, of course, as they had no control at all—but they reported a significantly enhanced *impression* of such control.

Another study in this paradigm tested the effects of priority. For this study, instructions were given a few seconds before each movement, just prior to movement, or after each movement had occurred. Participants felt decreased control over the arms' motions with late instructions, whereas the slightly early and just-in-time instructions yielded similarly enhanced experiences of agency as compared to a no-instruction group. These results suggest that even a minor shift in timing—in which an instruction appears just a moment too late—irremediably undermines the illusion of agency that the instruction provides. Knowing what another person's arms are doing after they've finished their motion produces no notable increment in the feeling of vicarious agency for that motion.

The priority effect seems well illustrated in the feeling of uncanny agency that comes when we serendipitously anticipate an event. Thinking about a friend just before the friend calls on the phone, for example, prompts an odd sense of agency—we feel as though we'd conjured them up (Blackmore, Galaud, & Walker, 1994). Experiences like this one are so profound that they often are reported as the front line of evidence among believers in supernatural phenomena ("ESP must exist—remember that time I thought of Aunt Milly from Idaho just before she called?"). Thinking of the caller afterward, of course, would be entirely unimpressive, as the absence of proper priority would undermine any sense of personal agency. The matter of timing is crucial in the perception of willed action, so crucial that even happenstance events may be perceived as under one's control when they occur just after one has happened to think about them.

Exclusivity Effects

The third principle of apparent mental causation is that people see their thoughts as causing events to the degree that there are no other plausible candidate causes. When the thought and only the thought precedes an event—no one else is thinking something similar, for example, or doing something that seems influential—the person will experience the event as flowing from that thought.

We are often quite sensitive to the possibility that there are causes beyond our own thoughts that might produce our action. When we are thinking of having the halibut in a restaurant, for example, and someone else at our table orders the halibut just before we've been able to say that's what we want, we can feel "scooped." We may think that we should pick something else so it doesn't look as though we were merely copying our fellow diner. This sensitivity suggests a more general readiness to perceive that our conscious will is challenged by external authors. We become ill at ease when we are faced with questions of our own free will not so much because we have some aversion to causal determinism arising within us, but because of the concern that we are being pushed around or influenced by others (Bargh, this volume; Wegner & Sparrow, 2004). It is curious that we humans have bodies and minds so well constructed for proprioception—we can perceive our actions through intricate pathways of muscle sense, vision, joint and tendon movement sensations, vestibular senses, and more (e.g., Jones, 1988)—but we nonetheless discount all this internal evidence of our own causal influence when other people might be causing our actions. All it seems to take is another's prior movement, or even the hint of a command, and we relinquish much of our own experience of will and allocate responsibility to the other.

A renowned instance of this effect occurred in the obedience studies conducted by Milgram (1963). Research participants were led to believe that they were teaching another participant in an experiment by applying electrical shocks whenever he performed incorrectly, and many were found to apply such shocks willingly—to the point of apparently placing him in grave danger and possibly causing his death. Yet these people were willing to accept only a modicum of responsibility for this action. Participants obeying the experimenter reported what Milgram called an *agentic shift:* "[T]he person entering an authority system no longer views himself as acting out of his own purposes but rather comes to see himself as an agent for executing the wishes of another person" (Milgram, 1974, p. 133).

Exclusivity effects on the experience of will have been observed in studies of sensitivity to fine differences in the timing of action and gaze (Sparrow & Wegner, 2006). For these studies, a participant was asked to tap out the letters of the alphabet in order with a conductor's baton by following a line connecting letters on a maze, each on the click of a metronome. This was entirely straightforward for participants, as all were fairly familiar with the alphabet. After each completion of the maze, participants took a minute to rate the action on a set of scales measuring their experience of authorship. The study found that the experience of will was reduced when the experimenter pointed or merely gazed at the alphabet letters one in advance of the participant's current letter—and that the experience of will was enhanced when the experimenter pointed or gazed at the letters one behind the participant's current letter. Apparently, it doesn't take much in the way of social circumstances to override the sources of authorship information in body and mind to produce an alteration in overall experience of will. Someone else doing an action just before or after we do it makes us feel differently about whether we did it.

Drawing causal inferences about our own thoughts, then, is a major way in which we develop experiences of willing what we do. The sense of magic in the self is produced by mental processes that perceive the consciously accessible parts of the action puzzle—the thoughts about the action that come to mind, and the perceptions of the action itself. Sometimes these processes have access to yet other sources of evidence, as when people use perceptions of their own effort to draw inferences about their authorship of action (Preston & Wegner, 2007). But experience of apparent mental causation renders the self magical because it does not draw on all the evidence. We don't have access to the myriad neural, cognitive, dispositional, biological, or social causes that have contributed to the action—nor do we have access to the similar array of causes that underlie the production of the thoughts we have about the action. Instead, we look at the two items our magic selves render visible to us—our conscious thought and our conscious perception of our act—and believe that these are magically connected by our will. In making this link, we take a mental leap over the demonstrable power of the unconscious to guide action (e.g., Bargh, 2004; Bargh & Barndollar, 1995; Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trotschel, 2001) and conclude that the conscious mind is the sole player. We discern our magical role in the world by reference to any apparent premeditation of the world's events that we can find in our own minds.

BREAKING THE SPELL: TRYING TO LOOK BEHIND THE CURTAIN

Does knowing how the trick works undo the magic? Many of the most strident arguments for free will hinge on the idea that a scientific understanding of human behavior could potentially ruin everything. The magic will be undone, the glorious human spirit will be cheapened, demystified, and rendered grotesque. We will uncover the trolls operating the machinery in the dungeon, and we will never again be able to appreciate the sparkling radiance of the

Magic Kingdom of the self. Or, more realistically, we will uncover the genetic codes that produce neural structures that allow incoming sensations produced by social and situational factors to contribute to the cognitive computations that incline our motor output processes to lead us to behave—and *then* we lose the magic. Now, of course, we tend to worry about the dangers of unweaving the rainbow in all of science (Dawkins, 1998)—but we most seem to fear the loss of magic that might ensue if we came to understand ourselves (Dennett, 1984, 2003; Greene & Cohen, 2004). When we know the trick of what makes us behave, the magic of free will could be lost.

Magic does seem a fragile thing in some cases. Consider the magic of love. When people in close relationships are asked to explain why they feel love toward their partner, they subsequently rate their love for their partner as less than if they were prompted to give no explanation at all (Seligman, Fazio, & Zanna, 1980). The love doesn't go away entirely, but it does decline a bit for the moment. This same effect occurs for other emotional states; a number of studies have revealed that explaining a feeling can have the effect of dissipating that feeling (Pennebaker, 1997; Wilson, Gilbert, & Centerbar, 2003). There is something about portraying deep and meaningful things in layers of analysis that makes their meaning less compelling. Just try to explain a joke to someone and watch the mirth drain from the room.

Explanations also seem to dismantle the magic of evil. One of the key fears of psychologists who try to understand and explain the behavior of criminals, psychopaths, or other villains, is that their explanations render the evil person into someone normal who is just behaving in a world that seems to require evil. Explaining evil, in other words, seems tantamount to condoning it (Miller, Gordon, & Buddie, 1999). To retain our full appreciation of evil—enough at any rate to work up an appropriate level of outrage and hatred for the evildoer—we almost need to *resist* understanding it. Roy Baumeister's (1997) superb book on the psychology of evil returns to this theme repeatedly, as he reveals the conflict we all feel in trying to establish a causal understanding of evil without simultaneously making it somehow less horrible and repugnant.

Generating explanations for the existence of our beliefs may also reduce their perceived value. Preston and Epley (2005) have found that when people give explanations for a belief, they feel the belief is less valuable. When people use the belief as an explanation for other things, however, they feel that the belief is more valuable. Certain beliefs, such as a belief in God or a belief in personal free will, are easy to use as explanations—we could spend all day talking about what God could do, or what a person with free will might choose. But these sorts of beliefs are themselves difficult to explain. Why does God do what He does, and why does free will opt for fried dumplings instead of mixed greens? The finding that people who are urged to explain beliefs feel the beliefs are less valuable, like the findings for love and for evil, reveal that the simple process of explaining may indeed dispel magic.

These studies suggest that it is possible to "break the spell"—given the right spell and the right explanatory counterspell. There do seem to be cases when thinking about how something works can reduce our experience of magic. But is this the natural result of psychological explanation? Must the self be destroyed by its own explanation? The fact is, there are many mysteries that do *not* lose their poetry merely because they have been solved. Do people actually fall out of love and part forever when they've paused to "count the ways"? Probably not. And understanding evil doesn't make us treat it much differently either. Many legal cases have been tried in which those accused of crimes have attempted to escape punishment with a "good explanation"—usually an insanity plea of some kind (Denno, 2002)—but aside from a few anomalies, these explanations typically do not shield the accused from punishment. Indeed, the forms of incarceration that have been invented for people who plead insanity are sometimes more odious than those for people who are merely found guilty (Monahan et al., 1995).

The magic of self does not readily go away when we explain action. In fact, the self seems remarkably resistant to reports of its demise, cropping up again and again in most every living human. One reason for this resiliency could be that the self is an ongoing experience (Wegner, 2004). Every time we think of doing something and then do it, we play the trick on our minds. It seems like magic each time. Even if we have somehow overcome the magic we experienced when we thought of getting a cup of coffee and found it in our hands—the next moment we are thinking of going to the window to check on the weather and there is magic once again! Perhaps the sheer frequency of our experiences of conscious agency is sufficient to overwhelm the nattering of our inner skeptic telling us our behaviors are caused by mechanisms of mind and not by our free willings. We could be convinced of the magic self by its mere doggedness.

But there is more. Just as a joke that is repeated again and again becomes less funny, there is probably nothing about the frequency of free willing alone that keeps the magic of self alive. Repetition doesn't build the illusion. Rather, the illusion of the magic self is *inherently persistent*. This is a trick that we can't see through, an illusion that cannot be spoiled by knowledge of how and why it happens. There are visual illusions that have this power—they continue to fool us even though we know they're illusions. For example, Roger Shepard (1990) drew tables that have this property (see figure). The table on the left looks longer than the one on the right—in fact, the two of them *seem* very different. Yet if you cut out one table top and lay it on the other, you will find that the surfaces coincide exactly. You probably shouldn't really cut them out or it will ruin this perfectly nice book—just take my word for it (and Shepard's), and truly believe in your heart that the two surfaces are identical. And despite this belief, you will find that each time you look at them, the one on the left looks longer.

The magic of self is just like this. It is not logical. It doesn't go away when you know how it works. It still feels as though you are doing things, freely willing them, no matter how much you study the mechanisms of your own

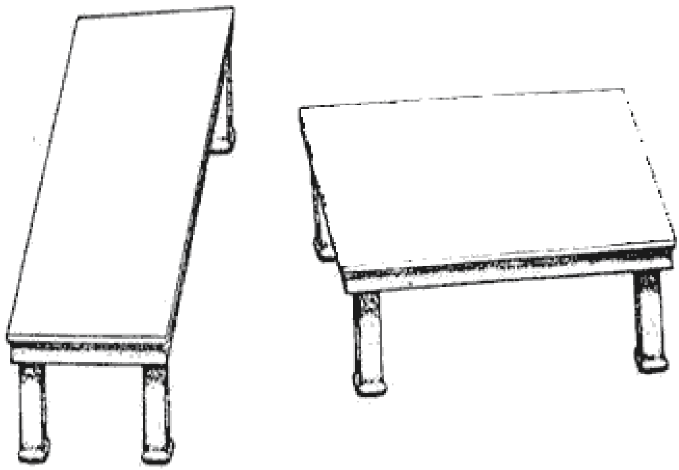

Figure 11.1.

behavior or gain psychological insight into how all people's behavior is caused. The illusion of self persists. This is why hand-wringing about the potential dangers of determinist thinking for morality (Shariff, Schooler, & Vohs, this volume) or for the law (Greene & Cohen, 2004) doesn't make a whole lot of sense. The agent self is an illusion, but it is not an illusion that is going to be whisked away by any amount of scientific explanation or heartfelt rhetoric. It is persistent. The magical self is here to stay.

I'm a case in point. I've devoted years of my life to the study of conscious will; I've written a book on how people experience an illusion of conscious will; I've taught seminars in the topic; and my lab has produced an array of experiments examining the wellsprings of the experience of will. If the illusion could be dispelled by explanation, I should be some kind of robot by now, a victim of my own nefarious schemes. No self, no magic, no inner agent. Yes it's true, when I'm on the dance floor I may look a bit robotic to some—but I'm happy to report that despite my personal flurry of illusion busting, I remain every bit as susceptible to the experience of conscious will as the next person. It feels like I'm doing things.

Why is the illusion of conscious will so persistent? One possibility is that the processes of causal inference we use to establish our feelings of will are not entirely logical. That is, information about the operation of causal consistency, priority, and exclusivity may not combine in a way that is simple or additive. The persistence of the illusion of conscious will suggests that we can experience this illusion at times even when we are faced with causal information that indicates we are *not* the authors of our action. The fact that people in our various experiments could accept authorship for events over which they clearly had no control—such as another person's movements, or an event happening at a distance—suggests that there is a magnetic quality to the experience of will, an attractiveness that draws us toward it despite clear evidence that it cannot be true.

Perhaps the best evidence of the illusion of agency swamping causal logic appears in a set of experiments conducted by Woolfolk, Doris, and Darley (2006). These studies examined the perception of others rather than self-perception, but they make the point nonetheless. For these experiments, people were asked to judge the moral responsibility of a person who committed a killing—he shot a man. Some participants were told that the shooter did so under overwhelming duress; either he had been forced to shoot while he himself was standing in the aim of a firing squad with automatic weapons, or he had been forced to shoot because he had been given a drug that rendered him utterly at the mercy of others' commands. And, as we would expect, these external forces were judged to lessen his moral responsibility for the shooting. In line with theories of discounting in causal attribution (Kelley, 1972; McClure, 1998), people saw these forces as reducing his moral culpability.

The intriguing finding of these studies was what happened when some people were told that he *meant* it. Some participants read that the shooter wanted to do it (because the victim had done harm to him in the past), whereas others were not so informed. The result? The shooter who wanted to do it was judged more responsible than the one who didn't—*even when he performed the act in the presence of overwhelming external causation (the firing squad or the drug) that clearly made him do it!* These findings suggest that people judge moral responsibility in a way that does not follow from a simple model of causal logic. It is as though we would hold the moon responsible for orbiting the earth if we could somehow discern that the moon wanted to do this—despite the clear operation of the laws of gravity and motion.

It may be that the illusion of conscious will is persistent because we honor so deeply what people mean to do that we readily overlook the causal forces that have impinged on them to force their action. In terms of the apparent mental causation theory, it may be that consistency information trumps exclusivity information. Merely having a thought or desire to perform an action is seen as the beginning of a magical connection from self to action regardless of what is going on in the world outside the self. The persistence of the illusion of conscious will could be due to processes of social evolution that have led us to a profound appreciation of what people think about what they do. We are drawn to the illusion of conscious will because we value so highly the cognitive previews of actions that our minds can provide for us.

THE WELLSPRINGS OF MAGIC: ILLUSION EVOLVES

If conscious will is such a mesmerizing personal illusion that it persists under every sort of explanatory insult, perhaps that's because it has been bred into us by an evolutionary process that has found the magic of the experience of will to be useful for the propagation of members of our species who fall under its spell. Perhaps people who don't experience conscious will fail to compete in society, fail to mate, or fail to parent successful offspring. Perhaps the societies they fashion work ineffectively and fail to survive. In contrast to Dennett's (2003, this volume) suggestion that freedom evolves, perhaps it is the illusion of freedom that evolves.

Why would an illusion evolve? The way to start thinking about this is to examine the downstream effects of the illusion. What personal or social consequences does the illusion of conscious will produce that might have proved sufficiently beneficial to individuals or to social systems that are advantageous to individuals that the whole mental apparatus for producing this illusion had to arise as an adaptation?

There are at least three such consequences of believing in the magic self, but deciding among them is difficult. As we well know, evolutionary arguments can be hard to test because they typically take the form of after-the-fact interpretations: How did a current trait of a living organism arise as an evolutionary adaptation to a particular environment by an ancestor organism that didn't have that trait? There are ways to test these stories (Cleland, 2001; Conway & Schaller, 2002), but the first step is telling the stories. So, in this final section of the chapter, let's consider three possible avenues for the evolution of magic: *social signaling*, *social task allocation*, and *social control*.

Social Signaling

As a first step in the functional analysis of conscious will, we need to pare it down to its basic element: The experience of willing an action is a conscious indication that one is the cause of the act. The experience is helpful as a marker of authorship, what I've called an "authorship emotion" (Wegner, 2002). Like the anger that translates an event of losing something into an experience that accentuates the loss in one's mind, conscious will translates an event of doing something into an experience that accentuates one's likely causal role in mind. Regardless of whether the experience of will is technically correct or not, it highlights in one's own mind the events in the world that seem to have been authored by self. The further usefulness of the experience of conscious will is that it gives us something we can communicate to others—a feeling of doing that we can then use to tell the world what we believe we have done.

The ability to give these self-assessments does not come without cost. The experience of willing is an *addition* to the usual processes that create action—a lean-to built beside the main barn where the actions are made. This add-on is an authorship module of mind that visits experiences of authorship on some of our behaviors, and fails to produce this experience for others. Many of the things we do might well occur without experiences of authorship, and in fact, this is true for a wide range of behaviors commonly described as automatic (Wegner, 2005). They simply happen, and we don't pay much attention to the fact that they issued from us. Indeed, we may not even notice who did it when an itch gets scratched or our position gets shifted in our chair. The authorship of such actions escapes our attention, and a range of such habits and rituals occur without the benefit of mental processes accompanying them that keep track of who did it. It is in the case of actions we feel will have freely willed that we have an experience that reminds us that we did them—and so tells us that these particular movements were not the result of external events, the movements of others' bodies, or the machinations of others' minds.

A key element needed for the feeling of will is the ability to think about our actions, ideally before they happen. This human capacity to experience mental *previews* of what we will do is, of course, the foundation of the common sense theory of intention—the idea that our thoughts truly cause our actions. But if we set aside this common sense theory in favor of the theory of apparent mental causation, why then would people need previews? If thoughts are not really needed prior to actions, why would evolution go to the extensive trouble to provide them for us? It could be that these thoughts arise not to cause action, but to signal its possible occurrence to us—a kind of warning that tells us what we can expect our bodies to do. Our thoughts about what we do may be part of a system for self-prediction. It would be impossible to tell a self-predicting system from an intending system if the self-predictions were accurate enough.

Self-predictions could be useful, much as it is useful to have dashboard gauges that tell us when the fuel is empty, or oven temperature lights to tell us when the turkey is roasting. The ability to think and talk about our actions well in advance of their occurrence is of particular utility for social purposes: We can tell when someone else might do something bad or good for us because their minds have handy self-prediction functions that have prompted them to tell us what they intend. When that angry fellow at the bar says he's going to break a pool cue over your head, you have a signal indicating what might happen. The interesting feature of such signals is that they can save *both* of you the trouble of actual physical harm. Communications of intention serve the purpose of making many potentially costly social actions unnecessary because the statement itself causes preventive responses.

Darwin (1872) pointed out that displays of dominance and submission are very useful to animals because they regularly take the place of actual conflict. In humans, as in dogs, exhibitions of aggressive intention can trigger us to roll over on our back and expose our tummies in a gesture of good will and desire for scratching—and thus prevent the mayhem and mutual danger that might have ensued. Perhaps previews of our actions come to mind so we can convey our likely behavior to others before it happens, and so signal our way out of social emergencies before they occur. Conscious will and intention may be much like turn signals on automobiles, features that have arisen to prevent accidents and so save the car for another trip tomorrow.

Social Task Allocation

What else does conscious will do? Another evolutionary story could be told about the role that our self-knowledge of action tendencies can play in helping us to choose tasks that will be useful to us, and that will also be useful to society and so bring us social rewards such as paychecks and opportunities to mate. Finding the right niche for our own special talents may be facilitated by experiences of conscious will.

In many social animals, particularly social insects, the job a given organism will do in the society is determined by its inherited morphology. The phenotype of a given caste of ant, for example, may be that it has an unusually large head. This makes ants of this form very useful to the whole colony as specialized soldiers whose job is to act as doors, so ants of this type spend all their time around the entrances, using their heads to shut the passageway to intruders and opening the way only when ants of their colony prompt them by tapping antennae on their big noggins (Holldobler & Wilson, 1990). We humans have not evolved such a system for task allocation in our species, so even those with very large heads seldom get work as doors. Instead, we develop systems for task allocation determined in part by self-assessed expertise. People who discern that they have certain abilities often take roles in society that take advantage of those abilities. If a person has regularly experienced a sense of willing associated with particular actions—hitting a baseball, for example, or arguing a point—that person may well become a ball player or an attorney as a result of the self-knowledge that this experience has provided. Conscious will conveys the sense that "I can do this" and human social organization has use for people who can select tasks they can do, even if that work involves no more than blocking doors.

Much of the psychological literature on perceived control has focused on this idea—that it is good to perceive control over those things one does indeed control, and not good to perceive control over those things that are in fact out of hand (Baumeister, Heatherton, & Tice, 1995; Burger, 1989; Folkman, 1984; Haidt & Rodin, 1999; Peterson, Maier, & Seligman, 1993). The experience of conscious will provides an anchor of sorts, an internal point of reference that is the mind's "best estimate" of whether the event in question might indeed be traceable to oneself. Although this estimate can never be correct scientifically in all respects (Davies, 2007), it certainly has its uses. The magic self is a natural guide to the roles one can play most effectively in life and in society. It tells us what we can and cannot do.

Social Control

Another social function for conscious will is to ready individuals to accept responsibility. As the theory of apparent mental causation suggests, the feeling of willing may be a poor indication of true causal responsibility, as this feeling can come and go in error. However, the theory also suggests that the feeling of willing that does arise in an individual for any action will compel that individual to accept personal responsibility for that action. Right or wrong, such responsibility acceptance then prepares the person for the experience of moral emotions such as pride for right action and guilt for wrong action. The experience of conscious will provides a unique inner signal, a first-person experience of responsibility that makes the person "own" morally relevant actions.

There are many instances when people don't accept complicity in actions that have moral overtones. Sometimes people are deceptive about these things—they deny performing crimes or they claim good deeds as their own—but the responsibility acceptance fostered by apparent mental causation guarantees that they will not be universally deceptive. The feeling of conscious will resonates in the person's mind even when there are clear reasons to lie about authorship of an action, and the feeling may guide the person to admit to wrongdoing (or own up to the lack of right-doing) even in the face of these reasons. This inner feeling of doing can behave like a "conscience," weighing in to make the person truthful about moral actions. Even if that truth is the expression of an illusion, it is an illusion that derives from the person's own best guess about the authorship of the action.

The acceptance of personal responsibility is a useful step toward successful social control of individual moral action. When society delivers its third-person judgments of responsibility—as when the law says someone is a criminal, or when a parent praises a child for helping with dinner—the individual's first-person feelings of responsibility will incline the individual to comply with these external judgments of culpability. Being sent to jail, or even being given a humanitarian award, would be difficult if we didn't authentically feel that we were the ones who had authored the moral action that had earned us those desserts. Quite simply, our inner feelings of doing give some license to the social world to hold us responsible.

Commentators who worry about the fate of legal and moral responsibility in a world that recognizes the illusory nature of conscious will have not come to appreciate the profound impact of the person's own moral sense. It matters far less for moral purposes what a person really did than what that person feels responsible for doing, and it is the feeling of responsibility that thus must be cultivated by social evolution. Our authorship processing modules may be good enough, often enough, that we typically get our responsibility roughly right. Like bonobos who hold each other responsible for food theft, or who know who should be punished for free riding (Boehm, 2001), humans who have a ready sense of their own complicity in right and wrong actions are likely to work effectively in social settings and survive some trials of social evolution.

The acceptance of individual responsibility for moral actions is essential for the exercise of social control. Indeed, the actions for which people experience free will are typically those actions that are most likely to be susceptible to modification by social consequences. In the study of animals, voluntariness is usually defined in terms of behavior modifiability (Passingham, 1993). Likewise in humans, actions we perceive as voluntary are also actions that are susceptible to modification through reinforcement. The things people feel apparent mental causation for, then, are those that are likely to be the focus of attempts at social control. If you say you can postpone a sneeze, people may pay you to do it when they want quiet—whereas if you find the sneeze inevitable, few attempts at control will come your way. This is nothing to sneeze at, however, as the identification of behavioral candidates for social control is serious business when it comes to criminal or immoral behavior. Far from eliminating responsibility for our behavior, then, the mental processes that produce the illusion of conscious will seem to be part of the mechanism that creates such responsibility and makes behavior more open to modification.

CONCLUSION: LIVING WITH AN ILLUSION

The life of the magician is not easy. Like Harry Potter, each of us must make sense of our amazing tricks of action and somehow fit our understanding into a sensible view of the world. This elliptical-peg-into-trapezoidal-hole problem is not an easy fit, and anguish surrounding the issue of free will and determinism echoes throughout philosophy and psychology. The chorus of discordant voices in this volume reveals that psychology continues to struggle with this conundrum. The uneasy solution suggested in this chapter involves learning to live with the magic.

The solution begins with recognizing that the magic is a little show we put on for ourselves. The sense of what we consciously will is only part of an authorship estimation system of mind, which can thus be mistaken—and a number of experiments in my lab and others suggest that the experience of will can not only be off, it can be wildly wrong. The second step toward solving the problem of free will is recognizing that, even when it is wrong, the magic we perceive in ourselves trumps other explanations *in our own minds.* Like that visual illusion of disparate tables that just won't quit even when we know it is an illusion, the magic self rules our intuitions and won't be undermined by analyses of its workings. (This realization is heartening for those worried about what would be left if the magic were gone: The magic is here to stay.) The third step in solving the problem of free will is discerning what functions this experience served in our biological and social evolution—and which of these might be so crucial that we absolutely had to have this magic installed in our heads. The third step is just beginning, as the understanding of the evolution of mind will take time. Until we take that step, we should be content to continue conducting scientific psychology to understand the mind, secure in the realization that the scientific discovery of our inner processes will never make us any less magical to ourselves.

REFERENCES

Aarts, H., Custers, R., & Wegner, D. M. (2004). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition, 14*, 439–58.

Aarts, H., Wegner, D. M., & Dijksterhuis, A. (2006). On the feeling of doing things: Dysphoria and the implicit modulation of authorship ascription. *Behaviour Research & Therapy, 44*, 1621–27.

Angel, L. (1989). *How to build a conscious machine.* Boulder, CO: Westview Press.

Bargh, J. A. (2004). Bypassing the will: Towards demystifying the nonconscious control of social behavior. In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious.* New York: Oxford University Press.

Bargh, J. A. (this volume). Free Will Is Un-natural. In J. Baer, J. C. Kaufman & R. F. Baumeister (Eds.), *Are We Free?* New York: Oxford University Press.

Bargh, J. A., & Barndollar, K. (1995). Automaticity in action: The unconscious as repository of chronic goals and motives. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 457–81). New York: Guilford.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trotschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81*, 1014–27.

Baumeister, R. F. (1997). *Evil: Inside human cruelty and violence.* New York: W. H. Freeman.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1995). *Losing control.* San Diego, CA: Academic Press.

Blackmore, S. J., Galaud, K., & Walker, C. (1994). Psychic experiences as illusions of causality. In E. Cook & D. Delanoy (Eds.), *Research in parapsychology 1991* (pp. 89–93). Metuchen, NJ: Scarecrow.

Boehm, C. (2001). *Hierarchy in the forest: The evolution of egalitarian behavior.* Cambridge, MA: Harvard University Press.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology.* Cambridge, MA: MIT Press.

Brown, J. W. (1989). The nature of voluntary action. *Brain and Cognition, 10*, 105–20.

Burger, J. M. (1989). Negative reactions to increases in perceived personal control. *Journal of Personality and Social Psychology, 56*, 246–56.

Cannon, W. B. (1942). "Voodoo" death. *American Anthropologist, 44*, 182–90.

Claxton, G. (1999). Whodunnit? Unpicking the "seems" of free will. *Journal of Consciousness Studies, 6*, 99–113.

Cleland, C. E. (2001). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science, 69*, 474–96.

Conway, L. G., III, & Schaller, M. (2002). On the verifiability of evolutionary psychological theories: An analysis of the psychology of scientific persuasion. *Personality and Social Psychology Review, 6*, 152–66.

Darwin, C. (1872). *The expression of emotions in man and animals.* New York: D. Appleton.

Davies, P. S. (2007). What kind of agent are we? A naturalistic framework for the study of human agency. In D. Ross, D. Spurrett, H. Kincaid, & L. Stephens (Eds.), *Distributed cognition and the will.* (pp. 39–60) Cambridge, MA: MIT Press.

Dawkins, R. (1998). *Unweaving the rainbow.* New York: Houghton Mifflin.

Dennett, D. (1984). *Elbow room: The varieties of free will worth wanting.* Cambridge, MA: MIT Press.

Dennett, D. (2003). *Freedom evolves.* New York: Viking Press.

Dennett, D. (this volume). Some observations on the psychology of thinking about free will. In J. Baer, J. C. Kaufman & R. F. Baumeister (Eds.), *Are We Free?* New York: Oxford University Press.

Dennett, D. C. (1994). The practical requirements for making a conscious robot. *Philosophical Transactions of the Royal Society, A349*, 133–46.

Denno, D. W. (2002). Crime and consciousness: Science and involuntary acts. *Minnesota Law Review, 87*, 269–400.

Folkman, S. (1984). Personal control and stress and coping processes: A theoretical analysis. *Journal of Personality and Social Psychology, 46*, 839–52.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions: Biological Sciences, 359* (1451), 1775–85.

Haidt, J., & Rodin, J. (1999). Control and efficacy as interdisciplinary bridges. *Review of General Psychology, 3*, 317–37.

Holland, O. (Ed.). (2003). *Machine consciousness.* Exeter, UK: Imprint Academic.

Holldobler, B., & Wilson, E. O. (1990). *Ants.* Cambridge, MA: Harvard University Press.

Hume, D. (1888). *A treatise of human nature.* London: Oxford University Press. (Original work published 1739)

Jones, L. A. (1988). Motor illusions: What do they reveal about proprioception? *Psychological Bulletin, 103*, 72–86.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151–74). Morristown, NJ: General Learning Press.

Kelley, H. H. (1980). Magic tricks: The management of causal attributions. In D. Gurlitz (Ed.), *Perspectives on attribution research and theory: The Bielefeld Symposium* (pp. 19–35). Cambridge, MA: Ballinger.

Lynn, S. J., Rhue, J. W., & Weekes, J. R. (1990). Hypnotic involuntariness: A social cognitive analysis. *Psychological Review, 97*, 169–84.

Lynn, S. J., Weekes, J. R., Matyi, C. L., & Neufeld, V. (1988). Direct versus indirect suggestions, archaic involvement, and hypnotic experience. *Journal of Abnormal Psychology, 97*, 296–301.

McClure, J. (1998). Discounting causes of behavior: Are two reasons better than one? *Journal of Personality & Social Psychology, 74*, 7–20.

Michotte, A. (1963). *The perception of causality* (T. R. Miles & E. Miles, Trans.). New York: Basic Books.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–78.

Milgram, S. (1974). *Obedience to authority.* New York: Harper & Row.

Miller, A. G., Gordon, A. K., & Buddie, A. M. (1999). Accounting for evil and cruelty: Is to explain to condone? *Personality & Social Psychology Review, 3*, 254–68.

Monahan, J., Hoge, S. K., Lidz, C., Roth, L. H., Bennett, N., Gardner, W., et al. (1995). Coercion and commitment: Understanding involuntary mental hospital admission. *International Journal of Law & Psychiatry, 18*, 249–63.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–59.

Passingham, R. E. (1993). *The frontal lobes and voluntary action.* Oxford, England: Oxford University Press.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8* (3), 162–66.

Peterson, C., Maier, S. F., & Seligman, M. E. P. (1993). *Learned helplessness: A theory for the age of personal control.* New York: Oxford University Press.

Preston, J., & Epley, N. (2005). Explanations versus applications: The explanatory power of valuable beliefs. *Psychological Science, 16*, 826–32.

Preston, J., & Wegner, D. M. (2007). The eureka error: Inadvertent plagiarism by misattribution of effort. *Journal of Personality and Social Psychology, 92*, 575–84.

Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: The role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology, 91*, 218–31.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots, 12*, 13–24.

Seligman, C., Fazio, R. H., & Zanna, M. (1980). Effects of salience of extrinsic rewards on liking and loving. *Journal of Personality and Social Psychology, 38*, 453–60.

Shariff, A. F., Schooler, J. W., & Vohs, K. D. (this volume). The hazards of claiming to have solved the hard problem of free will. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are We Free?* New York: Oxford University Press.

Shepard, R. N. (1990). *Mindsights.* New York: W. H. Freeman.

Sparrow, B., & Wegner, D. M. (2006). *The experience of authorship in coaction.* Unpublished manuscript, Cambridge, MA.

Spence, S. A. (1996). Free will in the light of neuropsychiatry. *Philosophy, Psychiatry, & Psychology, 3*, 75–90.

Thompson, S. C., Armstrong, W., & Thomas, C. (1998). Illusions of control, underestimations, and accuracy: A control heuristic explanation. *Psychological Bulletin, 123*, 143–61.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences, 7*, 65–69.

Wegner, D. M. (2004). Précis of *The Illusion of Conscious Will. Behavioral & Brain Sciences, 27*, 649–92.

Wegner, D. M. (2005). Who is the controller of controlled processes? In R. Hassin, J. Uleman, & J. Bargh, A. (Eds.), *The new unconscious* (2nd ed., pp. 19–36). New York: Oxford University Press.

Wegner, D. M., & Erskine, J. (2003). Voluntary involuntariness: Thought suppression and the regulation of the experience of will. *Consciousness & Cognition, 12*, 684–94.

Wegner, D. M., & Sparrow, B. (2004). Authorship processing. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (3rd ed., pp. 1201–9). Cambridge, MA: MIT Press.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology.*

Wegner, D. M., & Wheatley, T. P. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*, 480–92.

Wilson, T. D., Gilbert, D. T., & Centerbar, D. B. (2003). Making sense: The causes of emotional evanescence. In I. Brocas & J. Carrillo (Eds.), *The psychology of economic decisions: Vol. 1. Rationality and well-being.* New York: Oxford University Press.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition, 100*, 283–301.

12 Some Observations on the Psychology of Thinking About Free Will

Daniel C. Dennett

Do we have free will, or don't we? The problem of free will is hard, and important. Indeed, I will argue that it is one of our hardest intellectual problems, and it is hard *because* it is also one of the most important intellectual problems we face. It is our tacit or subliminal recognition of the importance of the problem that makes it so difficult.

In my graduate student days, it fell to me to forge a signature on a legal document of some significance. If you must know, a friend asked me to intercept and deposit his fellowship stipend while he was away from Oxford during the long Christmas break. In order to prevent just such junkets, the fellowship authorities included with the stipend a receipt that had to be signed and returned immediately, so they could check the dates and postmarks—and signatures. Failure to be present and accounted for when the check arrived was grounds for termination of the fellowship. Before departing for the south of France, my friend gave me several samples of his signature, which I practiced diligently, writing it well over a hundred times, until I had it just about perfect. When the check arrived at his flat, I practiced the signature another fifty times or so, then put the fatal document in front of me and proceeded to scrawl the shakiest, least convincing version of it I had ever written. My hand was trembling, my pulse was racing, and I learned at that moment that a life of crime was not for me. Fortunately, the abominable signature survived the scrutiny of the fellowship overseers so no horrible consequences ensued. In the anxious aftermath, a much better strategy occurred to me. I should have asked my wife to embed the official document somewhere in a largish sheaf of papers to be signed; I could have signed them all with *élan* not knowing which was the only one that mattered (an old device, related to the practice of putting blanks in the rifles of many of the members of a firing squad. That way nobody knows whether they are committing homicide.)

I was reminded of this embarrassing episode in my past when I recently confronted the question of why so many really intelligent people write such ill-considered stuff when the topic is free will. The answer, as I will try to explain, is that in some inchoate way they sense—correctly—that it really matters, and they just don't want to contemplate the implications straightforwardly, in case the truth is too horrible to live with. This makes wishful thinking and other distortions of reason almost irresistible. If the arguments they are tempted by were somehow embedded in less forbidding contexts, they would see through them in an instant. People don't do their best work when they think the stakes are astronomically high. Some of them may even be seduced by the following quite reasonable consideration: When we consider whether free will is an illusion or reality, we are looking into an abyss. What *seems* to confront us is a plunge into nihilism and despair. Our whole reason for living is jeopardized. What to do? If it is really as important as all that, perhaps what would be rational to do is *blow more smoke*. Whatever you do, don't try to get clear about this! Don't let the cat out of the bag. And then there are those who see the smoke screen for what it is but then mistakenly dismiss the problem, which really is important. I will be concentrating on this dismissal here.

In 1984 I published a small book about free will, *Elbow Room: The Varieties of Free Will Worth Wanting*. In it I tried to show that these fears are reasonable but mistaken: We *can* have all the varieties of free will *worth wanting*. Philosophers have managed to define some varieties of free will that are indeed incompatible with what we think we know about the physical world from science, but these varieties, it turns out, are negligible. Nobody needs to lose any sleep over the fact that they are beyond our reach. The considerations I advanced for this optimistic conclusion were, I think, well grounded, but whatever the merits of my case, there were two points of concern on which I could not yet deliver all the goods. First, I needed a more detailed naturalistic theory of consciousness, since many people share the intuition of philosopher P. F. Strawson that genuine freedom depends on an agent's behavior being "intelligible in terms of conscious purposes rather than in terms only of unconscious purposes" (1962, pp. 9–91, quoted in Dennett, 1984, pp. 36–37). Second, I needed a more foundational account of evolution by natural selection, since I was relying on evolutionary theory to provide the *design work* that, I claimed, distinguished genuinely free agents from less sophisticated (and hence morally incompetent) agents.

I set out to fill these gaps, in *Consciousness Explained* (Dennett, 1991) and *Darwin's Dangerous Idea* (Dennett, 1995), and my conviction that I was on the right track was bolstered by a curious pattern I observed in the critical reactions to my uncompromising materialism in both these books: My critics would begin with one technical challenge or another ("But what about this:...?") and after I'd parried their point, they would come up with another, and perhaps a third or fourth, but eventually, after I had responded to their apparent satisfaction to their technical objections, they would say, in one way or another, "Very well. But *what about free will?*" This was the hidden agenda that was driving their skepticism all along: the concern that if, as I was arguing, consciousness could be explained as a material phenomenon, and evolution could explain how it, and all the competences associated with it, came to be, the resulting picture of mankind would somehow fall short of providing us with enough...*magic* to give us the free will we desperately want to believe we have. I use the term "magic" advisedly, inspired by a passage in Lee Siegel's excellent book on Indian street magic, *Net of Magic: Wonders and Deceptions in India* (1991):

> "I'm writing a book on magic," I explain, and I'm asked, "Real magic?" By *real magic* people mean miracles, thaumaturgical acts, and supernatural powers. "No," I answer: "Conjuring tricks, not real magic." *Real magic,* in other words, refers to the magic that is not real, while the magic that is real, that can actually be done, is *not real magic.* (p. 425)

It can't be *real* if its explicable as a phenomenon achieved by a bag of cheap tricks. That is just what many people claim about consciousness.

Indeed, so powerful is that presupposition that there is a standard joke about my attempt to do just that: Dennett should have titled his book *Consciousness Explained Away.* It is also what many say about free will, and I'm sure this is no coincidence. So I decided I had to tackle the problem of free will again, armed with my theory of consciousness and my survey of the power of natural selection as a designing process. In my second book on free will, I tried to show that the varieties of free will worth wanting could indeed be composed of a bag of natural tricks, products of genetic and cultural evolution. According to *Freedom Evolves* (2003a), it is evolutionary biology, not (indeterministic?) physics, that accounts for free will. (A billion years ago, there was no free will on this planet, but now there is. The physics hasn't changed; the improvements in *can do* over the years had to evolve.) Once again, however, a chief source of resistance came from those who were reluctant to let go of a traditional, absolutistic variety of free will.

Consider these passages from two reviews: "One wants to be what tradition has it that Eve was when she bit the apple. Perfectly free to do otherwise. So perfectly free, in fact, that even God couldn't tell which way she'd jump" (Fodor, 2003). In other words, "one wants" a miracle, "the kind of absolute free will and moral responsibility that most people want to believe in and do believe in," as Galen Strawson (son of P. F. Strawson) said in his 2003 review. As he went on to say, that miraculous kind of free will can't be established by a materialist. Both Fodor and Strawson insist that this is what people want, and I have to agree with them that many people do want this, but what makes them so sure people are *right* to want this? Might they be lulled or gulled by a tradition of dualism that doesn't so much *solve* the free will problem as *hide* the problem behind an impenetrable shield of mysterious stuff or *postpone* the problem indefinitely? ("It's anybody's guess—and not the business of science—how mind-stuff manages to generate genuinely free decisions and get them implemented by a material body. Don't ask; don't tell.")

Is free will truly incompatible with materialism—deterministic or even just mechanistic materialism? This question has been nagging ever since Darwin put forward his revolutionary theory of natural selection. Thomas Henry Huxley, known as "Darwin's bulldog" for his forthright championing of the theory of evolution by natural selection in its early days, wrote a popular piece "On the Hypothesis That Animals Are Automata, and Its History," in which he confronted this dire question head-on and tried to mollify the gloomleaders (see Dennett, 1984, p. 7):

> I venture to offer a few remarks for the calm consideration of thoughtful persons, untrammeled by foregone conclusions, unpledged to shore-up tottering dogmas, and anxious only to know the true bearings of the case.

and goes on to insist that

> We are conscious automata, endowed with free will in the only intelligible sense of that much-abused term—inasmuch as in many respects we are able to do as we like.... (Huxley, excerpted in Chalmers, 2002, p. 30)

We are able to do as we like, Huxley avers, in the sense that if we are not imprisoned or paralyzed we can act as we *choose*. But are we able to *choose* as we like—and for that matter, are we able to *like* as we like? Without these supplements, Huxley's assurances ring a bit hollow. Here is where evolution comes to our rescue with a second wave of design: cultural evolution. The combination of genetic and cultural evolution does provide *Homo sapiens*—and only *Homo sapiens*, so far, on this planet—with precisely those features. Thanks to our enculturation, we have been endowed with perspectives that enable us (and only us) to reflect indefinitely on whether our choices are well grounded, whether we *ought* to like what we find ourselves liking, and so forth. Even when we discover, as we sometimes do, that it is difficult or impossible for us to revise some of our likes and dislikes, at least we can inform ourselves of this, and think about ways of working around them.

It is this open-ended ability we have to deal informedly and constructively with our own grounds for, and habits of, choice that gives us a variety of free will that underwrites moral responsibility and that is inaccessible to the cleverest and most spontaneous animals.

One might think that this would be enough, that since this naturalistic variety of free will preserves and explains what really matters—our belief in our own moral responsibility, and thus the denial of nihilism—it deserves to be called *real* free will. But the tradition of "real magic" is still so strong in these discussions that many thinkers, *on both sides of the issue*, continue to muddy the waters, however inadvertently. On this occasion I am going to set aside the many tantrums thrown by those who insist on the existence of traditional (mysterious) free will because I and others have dealt with them at great length elsewhere. Instead, I will concentrate on the overreactions to all this desperate posturing by some of my favorite thinkers.

In 2002, my friend the psychologist Daniel Wegner published his provocatively titled book, *The Illusion of Conscious Will*. I had read the book in draft, and loved it—but objected to the title, which played into the hands of the "real magic" crowd. As I said in *Freedom Evolves*,

> Recall the myth of Cupid, who flutters about on his cherubic wings making people fall in love by shooting them with his little bow and arrow. This is such a lame cartoonists' convention that it's hard to believe that anybody ever took any version of it seriously. But we can pretend: Suppose that once upon a time there were people who believed that an invisible arrow from a flying god was a sort of inoculation that caused people to fall in love. And suppose some killjoy scientist then came along and showed them that this was simply not true: No such flying gods exist. "He's shown that nobody ever falls in love, not *really*. The idea of falling in love is just a nice—maybe even a necessary—fiction. It never happens." That is what some might say. Others, one hopes, would want to deny it: "No. Love is quite real, and so is falling in love. It just isn't what people used to think it is. It's just as good—maybe even better. True love doesn't involve any flying gods." The issue of free will is like this. If you are one of those who think that free will is only *really* free will if it springs from an immaterial soul that hovers happily in your brain, shooting arrows of decision into your motor cortex, then, given what *you* mean by free will, my view is that there is no free will at all. If, on the other hand, you think free will might be morally important without being supernatural, then my view is that free will is indeed real, but just not quite what you probably thought it was. (Dennett, 2003a, p. 222)

I saw Wegner as the killjoy scientist who shows that Cupid doesn't shoot arrows and then entitles his book *The Illusion of Romantic Love*. Wegner does go on to soften the blow by arguing that "conscious will may be an illusion, but responsible, moral action is quite real" (p. 224). Our disagreement was really a matter of expository tactics, not theory. Should one insist that free, conscious will is *real* without being magic, without being what people traditionally thought it was (my line)? Or should one concede that traditional free will is an illusion—but not to worry: Life still has meaning and people can and should be responsible

(Wegner's line)? The answer to this question is still not obvious, but Wegner was surprised and dismayed by some of the hostile reactions he provoked. One commentator was inspired to call him a "cryptobehaviorist" who provided "terrifying interpretations of his experiments" (Baars, 2004). Wegner has recounted to me a variety of other panic-ridden reactions to his work.

More recently, the World Question Center on edge.org mounted its 2006 question, "What is your dangerous idea?" and my friend Richard Dawkins dashed off—and later regretted sending and tried unsuccessfully to retract—a piece inspired by his friend John Cleese's hilarious scene in *Fawlty Towers* in which he beats his automobile, "punishing" it for its poor performance. The image is unforgettable, but the conclusion Dawkins was tempted to draw was a non sequitur indeed:

> Why do we vent such visceral hatred on child murderers, or on thuggish vandals, when we should simply regard them as faulty units that need fixing or replacing? Presumably because mental constructs like blame and responsibility, indeed evil and good, are built into our brains by millennia of Darwinian evolution. Assigning blame and responsibility is an aspect of the useful fiction of intentional agents that we construct in our brains as a means of short-cutting a truer analysis of what is going on in the world in which we have to live. My dangerous idea is that we shall eventually grow out of all this and even learn to laugh at it, just as we laugh at Basil Fawlty when he beats his car. But I fear it is unlikely that I shall ever reach that level of enlightenment. (Dawkins, 2006; note that this appeared on edge.org, but was not included, with the rest of the answers elicited in John Brockman, ed., *What Is Your Dangerous Idea?* 2006)

What Dawkins was overlooking was the prospect that there might be some stable—indeed homeostatically maintained—middle ground in between the saints (who never need punishing) and the "faulty units" who really are so disabled that it would be as pointless a travesty to punish them as to punish Fawlty's car. By his own admission, Dawkins can't quite accept his own conclusion—not, I am sure, because he is insufficiently "enlightened" but because he has quite properly failed to convince himself that "blame and responsibility, indeed evil and good" are *just* "useful fictions." Why would Dawkins, of all people, think that useful fictions were always something to outgrow? The selfish gene is a useful fiction—that is to say, it encapsulates in a well-nigh irreplaceable *idealization* a real pattern that is otherwise indescribable. Much the same can be said of the simplifications that we rely on when we hold people responsible or call an act good or evil.

Another dear friend, Susan Blackmore, winds up her book *The Meme Machine* (1999) with a ringing disagreement with me about free will: "Dennett (1984) has described many versions of the idea of free will and argues that some of them are worth wanting. Unlike Dennett I neither think the 'user illusion' is benign, nor do I want any version of free will that ascribes it to a self that does not exist" (p. 237). Her reasoning is clear, and worth quoting at length. I agree with most of it:

> Benjamin chose cornflakes this morning for breakfast. Why? . . . Memes and genes together produced this behaviour in this environment. If asked, Benjamin will say that he chose the cornflakes because he likes them, or that he made a conscious decision to eat them today. But this explanation adds nothing. It is just a story Benjamin tells after the fact.
>
> So does Benjamin have free will or not? The critical question to ask is who do you mean by Benjamin? If by "Benjamin" you mean a body and brain, then certainly Benjamin had a choice. Human beings make decisions all the time. . . . Is this sufficient for what we call free will?
>
> I think not, because at the heart of the concept of free will lies the idea that it must be Benjamin's conscious self who made the decision. When we think of free will we imagine that "I" have it, not that this whole conglomeration of body and brain has it. Free will is when "I" consciously, freely, and deliberately decide to do something, and do it. In other words "I" must be the agent for it to count as free will.
>
> But if the memetic view I have been proposing here is right, then this is nonsense, because the self that is supposed to have free will is just a story that forms part of a vast memeplex, and a false story at that. . . . There is no truth in the idea of an inner self inside my body that controls the body and is conscious. Since this is false, so is the idea of my conscious self having free will. (Blackmore, 1999, pp. 236–37)

With what do I disagree? I disagree with her acquiescence (along with Wegner and Dawkins) in the traditional concept of free will, with its "inner self." The basically Cartesian idea of the self or ego or *res cogitans* as the inner (conscious) agent is indeed a huge mistake, and therefore, as she says, any view of free will that depends on it is bankrupt. But why accept that this is "the heart" of the concept of free will? That is the concession that gives the game to the traditionalists, by letting their antiquated and now utterly unmotivated vision of the "seat" of free will capture the term. Well, what else might be the heart? Here's a suggestion: *Free will is whatever it is that gives us moral responsibility* (if we ever have it). And *that* turns out to be the very body and brain (plus friends and acquaintances, tools and crutches, . . .) that she says actually make the decisions. The key to understanding *real* free will is recognizing that it does not reside in some concentrated internal lump of specialness, but in the myriad relations and dispositions of an enculturated, socialized, interacting, acknowledging, human agent. Tradition makes the Cartesian mistake of packing all the power into the inner puppeteer who pulls the body's strings. When we banish this inner agent, distributing its tasks throughout not just the entire brain, but the body and the "surrounding" cultural storehouse—the memes, plus a little help from our (human) friends—we don't have to banish free will! We can see it as a phenomenon distributed in space and time as well. That was the point of my ironic formula, in *Elbow Room* (1984, p. 143), "If you make yourself really small, you can externalize virtually everything." Don't make yourself—your *self*—small; that's the Cartesian error; recognize that there is a nonmysterious, and valuable, concept of a self that can be large enough to take responsibility and act morally. (In this regard, see also my discussions of Wegner's residual Cartesianism in Dennett 2003a, 2003b, 2004.)

Dawkins, overreacting to the foolishness of the Cartesian vision of free will and its absolutist concepts of good and evil, and moral responsibility, imagines throwing out the notions of good and evil altogether. Wegner reassures his readers that "moral action is quite real" but doesn't try to say how this can be (if conscious will is an illusion). Like Wegner, Blackmore shrinks from abandoning the notion of moral responsibility altogether, and she ends her book with a brief and optimistic look at how the world might seem to a living human body that had abandoned the perspective of the "selfplex" altogether: "This lack of self-concern means that you (the physical person) are free to notice other people more. Compassion and empathy come naturally.... Perhaps the greater part of true morality is simply stopping all the harm that we normally do, rather than taking on any great and noble deeds; that is, the harm that comes from having a false sense of self" (1999, p. 246). Perhaps. And perhaps not. Is she going to jettison our systems of law and punishment? Is she going to abandon the social leverage by which we encourage people to *take* responsibility for their actions? Is she prepared to dismiss the distinction between honesty and cheating as just another myth fostered by the traditional concept of free will?

Vohs and Schooler (in press) have recently described some pioneering research on the actual effect of the expression of doctrines about free will on behavior. In their first experiment, one group of students was given a passage to read from Francis Crick's *The Astonishing Hypothesis* that "claimed that rational, thinking people (such as scientists) have long denounced the idea of free will, noting that it is instead a byproduct of the human mind" Notice how seamlessly this statement welds the idea of free will to the idea of a magical, dualistic kind of inner chooser. No mention is made of morality or responsibility; the only idea that "rational, thinking people" have denounced is the "idea of free will." Which idea of free will? The control group was given a passage from the same book on another topic. Participants in both groups were then given a task in which there was an opportunity to cheat—to get more money for their participation in the experiment than they would earn by abiding by the rules—and those who had read the passage denying the existence of free will cheated significantly more often than those in the control group. This is not a result that supports Blackmore's optimism. Vohs and Schooler have yet to do the balancing experiment in which one group reads an authoritative statement that assures them that science has shown that free will is real—just not what they might have thought it was. The passage I have suggested to Schooler would be

> Scientists have established that every single action and thought that you have is caused by the current state of your brain and body, which in turn is caused by the interaction between your current environment as you experience it, your whole life history, and, of course, your initial genetic endowment.
>
> Some are tempted to conclude from these facts that we don't really have free will, but this is a mistake. Free will in the sense that matters, in the sense that makes you responsible for your actions and that gives meaning to both your strivings and your regrets, is determined by *how* your brain deals with the reasons it finds for acting. Philosophers have established that you can still have free will and moral responsibility when the decisions your brain arrives at are *your* decisions, based on your very own reasoning and experience, not on any brainwashing or manipulation by others. If your brain is normal, it enables you to consider and reconsider your options and values indefinitely, and to reflect on what kind of a person you want to be, and since these reflections can lead to decisions and the decisions can lead to actions, you can be the author of your deeds, and hence have free will in a very important sense.
>
> Some people have diminished free will and responsibility through no fault of their own: their brains malfunction or they have been kept ignorant of the facts and values that a normal person knows full well, but those who are fortunate enough to have had a normal upbringing arrive at adulthood with all the free will necessary to be held accountable for their actions.

Would this, or a more effective version of it, produce a diminution of cheating compared to the control group? It will be very interesting to discover. What Vohs and Schooler, and a few other psychologists, are now embarking on is a new, empirical investigation of how *what you believe* about free will influences your behavior, a theme that I have been hammering on for decades (e.g., Dennett, 1984, 2003a). It follows directly from this that scientific investigations of free will have an environmental impact—in particular an impact on the *belief environment* (Dennett, forthcoming) that has to be taken into account. In particular, if the popularization of the results of this research leads to widespread misunderstanding of its import—which would be my interpretation of the effect Vohs and Schooler have uncovered—it could have a seriously detrimental social effect. This does not at all vindicate the traditionalists who have distorted thinking on free will for decades, but it does uncover a powerful, and not ignoble, motivation for their sometimes deliberate misrepresentation (Dennett, 2003a, forthcoming).

What are people afraid of? Perhaps they are afraid of the burgeoning of complicating conditions. This is not unreasonable. Once absolutism is abandoned, all manner of paths open up down which we might not want people walking! Consider a few of the (apparent) possibilities:

YES, we have free will, but only if we take Prozac (or some other drug) once a day for the rest of our lives.

YES, we have free will, but only if we can master the stunt of squinting whenever we feel tempted to look too closely at the machinery involved.

It's like golf: Consider the golf pro's advice to keep your head down until you have completed your swing:

> But how can this be good advice? The ball leaves the club head in mid-swing, and after it has begun its trajectory, nothing that happens on the tee can alter what trajectory. Isn't the attention to details of the swing that occur after the ball leaves the club just so much body English? Not necessarily. For it maybe that the only way to get the right thing to happen up to the moment of impact is to look ahead and fix a more distant goal, counting on one's efforts to satisfy that goal to produce bodily motions that traverse just the right space at just the right speed. One would be foolish indeed to disregard the pro's advice on the basis of the argument given above that it couldn't make any difference. It could make all the difference. Sometimes the only way to get what you really want is to try to do something else. (Dennett, 1984, p. 16)

What if the parallel, in free will, to keeping your head down (in golf), is believing in an afterlife? Or believing in the Old Testament God? Is *that* too steep a price to pay for free will? What if you're simply unable to muster the conviction? Have we lost our virginity for free will?

Robert Frank (1988, pp. 111–12) draws our attention to the

> practice whereby many affluent couples in New York City recruit governesses for their children. . . . Apparently experience has persuaded many New Yorkers that the local labor market is not a good place to recruit people who perform reliably without supervision.
>
> The solution many of these couples have adopted is to advertise for governesses in Salt Lake City newspapers. they have discovered that persons raised in the Mormon tradition are trustworthy to a degree that the average New Yorker is not.

There are other unsettling prospects to explore:

YES, we have free will, but not everybody does. In fact, roughly 10% of the adult "healthy" population achieves the intuitively reasonable level of moral competence free will demands. Most people are too ill-controlled to be held responsible, and ought to be kept in a state of permanent childhood, indulged as best we can manage so that they don't become too unruly.

> YES, we have free will—most of us—but each of us is a sort of checkerboard of moral competence and incompetence. Trust alcoholic Adam with the lives of your children—unless he's fallen off the wagon, which might happen at any time. Trust *idée fixe* Irene unless the topic is abortion; she just stops thinking when that issue is raised. Trust adults in general about cooperation and integrity so long as they haven't taken too many economics courses!

In fact, of course, we already know that there are grounds for such subversive abridgements of our brittle, absolutistic doctrines of free will. We already know, as Tori McGeer (personal correspondence) has put it, that free will is not something that comes for free. We know it because if we have any self-knowledge to speak of, we discover such weaknesses in ourselves. And we discover them in our loved ones, and—hardest of all to accept—we discover them in our enemies, whom we are loath to let off the hook. We ought to admit, up front, that one of our strongest unspoken motivations for upholding something close to the traditional concept of free will is our desire to see the world's villains "get what they deserve." And surely they *do* deserve our condemnation, our criticism, and—when we have a sound system of laws in place—punishment. A world without punishment is not a world any of us would want to live in.

We need to coordinate our investigations of the role of censure and punishment (and praise and reward) in society with our investigations of the complexities of human motivation, and the role of beliefs—and beliefs in beliefs (Dennett, 2006)—sin shaping the perspectives of ourselves and our fellow citizens. This is going to be a ticklish task, in which missteps might be painfully amplified. No wonder our hands shake when we get to work on it.

REFERENCES

Baars, B. (2004). Comments at the Association for the Scientific Study of Consciousness, 7th annual meeting, Memphis.

Blackmore, S. (1999). *The Meme Machine.* Oxford: Oxford University Press.

Chalmers, D. ed. (2002). *Philosophy of Mind:Classical and Contemporary Readings,* New York: Oxford University Press.

Dawkins, R. (2006). response to "What Is Your Dangerous Idea?" on edge.org. Also in J. Brockman, ed., *What Is Your Dangerous Idea?*

Dennett, D. C. (1984). *Elbow Room: the Varieties of Free Will Worth Wanting.* Cambridge, MA: MIT Press.

Dennett, D. C. (2003a). *Freedom Evolves.* New York: Viking Penguin.

Dennett, D. C. (2003b). "The Self as a Responding—and Responsible—Artifact," *Annals New York Academy of Sciences, 1001*, 39–50.

Dennett, D. C. (2003c). "Making Ourselves at Home in Our Machines: The Illusion of Conscious Will," (review of Wegner, 2002) in *Journal of Mathematical Psychology, 47*, 101–4.

Dennett, D. C. (2004). "Calling in the Cartesian loans" *Behavioral and Brain Sciences.*
Dennett, D. C. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Viking Penguin.
Dennett, D. C. (forthcoming). "How to Protect Human Dignity From Science," in a volume on human dignity, ed. A. Schulman, commissioned by the Presidents Council on Bioethics.
Fodor, J. (2003). Review of *Freedom Evolves, London Review of Books.* March 5.
Frank, R. (1988). *Passions Within Reason: The Strategic Role of the Emotion*. New York: Norton.
Siegel, L. (1991). *Net of Magic: Wonders and Deceptions in India*. Chicago: Chicago University Press.
Strawson, G. (2003)., Review of *Freedom Evolves, New York Times.* March 2.
Strawson, P. F. (1962). "Freedom and Resentment," *Proceedings of the British Academy, 48*, 1–25. Reprinted in Strawson, ed., 1968, *Studies in the Philosophy of Thought and Action*, Oxford: Oxford University Press.
Vohs, K. D., & Schooler, J. (in press) "The Value of Believing in Free Will:
Encouraging a Belief in Determinism Increases Cheating." *Psychological Science.*
Wegner, D. (2002). *The Illusion of Conscious Will,* Cambridge, MA: MIT Press.

13 Whose Will? How Free?

George S. Howard

The free will–determinism debate has raged for over 2000 years. Many seemingly intractable controversies (like free will–determinism) seem to be caused by asking wrongheaded questions. I'll give you two contemporary examples: First, are you a masculine or a feminine person? The answer to this question is easy for me—I'm both, high masculine and high feminine. I'm androgynous. It was wrongheaded for us to set masculine traits in opposition to feminine characteristics. In that case, in order to endorse feminine traits one must deny masculine characteristics. Forty years of sex-role orientation research has demonstrated why setting the wrong combatants in opposition to one another (i.e., masculine, feminine) can lead to nasty, self-inflicted wounds.

For my second example, I'll apply the same intellectual move to another hot button, contemporary (at least in the United States) debate: Are you pro-life or pro-choice with respect to abortion? This false dichotomy condemns one to oppose abortion in order to be for "life." What a ridiculous corner to allow oneself to be painted into. I'm both pro-choice and antiabortion (Howard, 1994). Setting these independent positions into a false oppositionality represents a wrongheaded conceptual move that leads to a heated and unsatisfying discussion—just as it did when we opposed masculinity to femininity.

Now, let's perform the same conceptual surgery to the free will–determinism conundrum. Free will respects the power of self-determination and is properly placed in opposition to mechanistic or nonagentic determination (by genetic, biological, social, intrapsychic, environmental, etc., causes). The second dimension is complete determinism versus complete acausality.

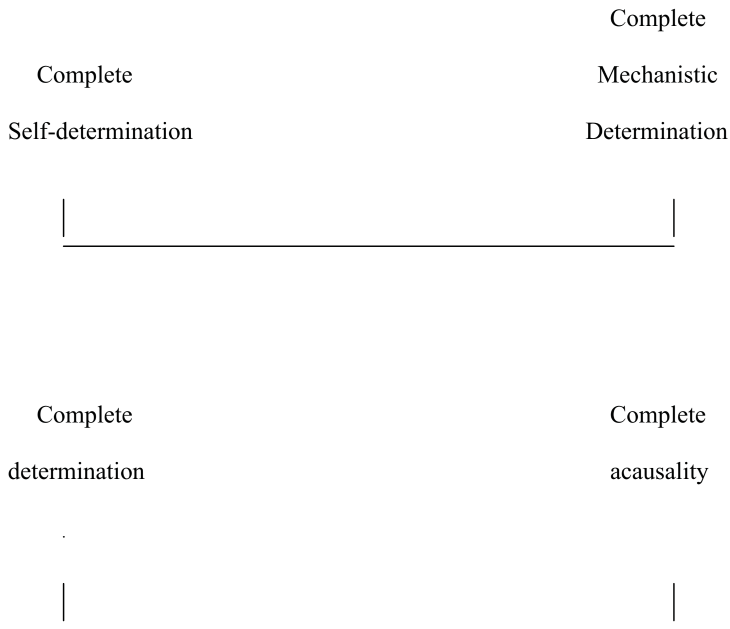

Figure 13.1.

This second dimension is easily dismissed. If you want to be a scientist, you had better be a determinist. Things are (and act) the way they are (and act) because something (or some things) caused them to be (or act) that way. It is a proper job for a scientist to find and document (via experimental studies) the cause–effect relations that form and guide human actions. Therefore, I am a determinist. But the first dichotomy in the accompanying figure asks me whether I believe in self-determination or mechanistic determination? Actually, I believe in both.

THE ROLE OF AGENCY IN A WORLD OF NONAGENTIC CAUSES

It is conceptually possible that humans have the complete power to behave in any way they choose. However, I know of no reputable psychologist who believes that genetics, biological factors, social structures, physical environments, reinforcement schedules, intrapsychic factors, and so forth, have absolutely no causal efficacy at all. I don't believe that extreme free will position either.

Conversely, although a few psychologists adopt the belief that all human action is completely mechanistically determined (e.g., Skinner, 1971), the relative importance of self-determination versus nonagentic, mechanistic causality in human action ought to be *empirically determined*. However, until recently, there was no way to precisely untangle the impact of self-determined actions from the ebb and flow of nonagentic causal influences in our lives. Would you like to hear how one can now perform such a tricky sequestering of agentic from nonagentic causes in an experiment?

The crucial aspect of the new methodology for investigating volitional behavior is that it requires the active cooperation of the research subject. If, for whatever reason, the subject chooses not to cooperate fully with the experimenter, a serious underestimate of that subject's degree of volitional (or self-determined) control will be obtained. One of the first procedures developed involves the experimenter dividing the total time for the experiment into a large number of equal-length time blocks. The experimenter then randomly assigns each of the time blocks to either "try to..." or "try not to..." conditions. For example, if one considered subjects' ability to control between-meals snacks, as a part of their ability to lose weight, the instructions for each subject on half the studies' days would be to "eat as many snacks as you wish" whereas on the other half of the days subjects would be instructed to "try not to eat any snacks." Differences in mean number of snacks consumed on "eat" versus "not eat" days is a reflection of the subject's ability to volitionally control snacking behavior.

The studies to be reviewed below probed subjects' capacity to volitionally control their actions. Is this teleological account the only interpretation that can be given for the data reported herein? Definitely not. One might choose to argue for a Humean-type, efficient cause explanation. For example, if subjects eat more peanuts on "try to eat" days than they do on "try not to eat" days, a critic might claim that such evidence does not imply volitional control of that action. Rather, the critic might assert that subjects have been socialized to play "good subject" roles in scientific studies. Thus, the difference in amount of peanuts eaten on "eat" versus "not eat" days is best attributed to conformity by the subject to the demands of the experiment, rather than being evidence of volitional control. (This conformity rival hypothesis is an example of a "variable correlated with the independent variable." This unfortunate coupling of volition with conformity is fundamental to this operationalization procedure and is not amenable to design control. Thus, the plausibility of this rival interpretation—that the mean difference between groups of days is caused by participants' conformity rather then their ability to self-determine—has to be specifically tested in several experiments to be described below.)

Although at first blush this might appear to be a serious criticism of this group of studies, closer inspection reveals that such interpretive difficulties are endemic to all research (although this fact is rarely acknowledged by psychological researchers). The problem is referred to by philosophers of science as the underdetermination of theory by evidence (Hansen, 1958; Kuhn, 1962, 1977). In its weakest form, the underdetermination thesis suggests that the meaning of a research finding is never transparent. Does this mean that there are no criteria whereby scientists can make objective (if fallible) judgments as to the probable meanings of research findings? Not at all. Cronbach (1982) points toward a partial solution to this dilemma by expanding the traditional notion of the concept of the "validity" of an experiment:

> Validity depends not only on the data collection and analysis *but also on the way a conclusion is stated and communicated*. Validity is subjective rather than objective: The plausibility of the conclusion is what counts. And plausibility, to twist a cliché, lies in the ear of the beholder.

The position espoused herein is that humans possess capacities that enable them to behave volitionally. If that position is correct, then individuals might be expected to be able to control their behavior in meaningful ways. But note that another investigator might see that same behavior as being under the control of stimuli, both internal and external. Such an individual might actually design the exact same studies as will be reviewed here (and, hopefully, would have obtained a similar set of results). But the account of the *meaning* of the findings would have been quite different—and perhaps equally plausible. No amount of evidence ever "proves" a theory, and relatedly, there can be multiple theoretical interpretations of any body of evidence. Hence, one is forced to a position like Cronbach's in which the scientific community evaluates the plausibility of various competing accounts of empirical findings. This chapter argues for the plausibility of a volitional account of a growing body of research evidence. However, other interpretations of these findings are also plausible.

The Evidence for Self-Determination

The caveat on the interpretation of research evidence aside, what are the data that suggest the importance of volitional control in human action? Using subjects' self-control as a warrant for volitional behavior, Howard and his colleagues (Howard, 1988, 1989; Howard & Conway, 1986; Howard, Curtin, & Johnson, 1988; Howard, DiGangi, & Johnson, 1988; Howard, Youngs, & Siatczynski, 1988; Lazarick, Fishbein, Loiello, & Howard, 1988; Steibe & Howard, 1986) conducted a series of studies that considered what proportion of a particular eating behavior was due to particular external, efficient cause influences, and what part was due to volitional control. Several studies dealt with subjects' ability to control eating peanuts. Peanut eating was chosen because it provides a noncontroversial dependent measure, and because eating peanuts is an activity subjects tend to enjoy but which they should be able to control. Here (Howard & Conway, 1986, Study 1) the effect size (partial Eta squared) for volitional control was 0.56, whereas the effect size for whether the food was kept in sight or out of sight was 0.13. Comparable figures for the second study were 0.57 for volition and 0.16

for whether the subject received a written reminder or not. Finally, in a third study (Howard, Youngs, & Siatczynski, 1988, Study 1), the effect size for volition was .53, whereas the effect size for a written reminder was 0.03. Therefore, in studies on the control of eating behavior, volition appears to be about five times more influential than certain (sight and reminder) external, efficient cause influences (i.e., the average effect size for volition was 0.56, whereas the average effect size for the efficient causes was 0.11).

Great care should be taken in interpreting the above findings. The major point is that we now are able to assess the influence of volition in eating in a rigorous, empirical manner. Beyond that, interpretation becomes difficult. For example, the ratio of the effect size of volition to the effect size of external factors should be viewed with extreme caution. Had we considered other and/or more nonvolitional factors in our studies, that ratio would likely have been reduced. Therefore, if one really were interested in volitional versus nonvolitional factors in eating behavior, these studies would represent but a first step toward developing a more complete understanding of the phenomenon that might be obtained by investigating additional external and/or organismic variables. Conversely, one might consider whether various self-control enhancing techniques might actually increase the volitional-to-nonvolitional ratio in our account of the phenomenon.

Have we tested the magnitude of the force of free will in these studies? Here we would respond with a resounding "No." Imagine a hypothetical subject, who was told on certain days to try to eat as few peanuts as possible (a "don't eat" day), who thought the following:

> I know I'm not supposed to eat today, and I could resist if I wanted to (Assume this is to be true for now). But I'm not that invested in the results of this study, and I've studied hard today, I feel I deserve a treat. I'm going to eat some nuts because I feel like it.

In such a case, the subject might have exercised his/her free will, but the act of eating the nuts would serve to *decrease* the volitional component of behavior observed in the study. Insofar as such decisions do actually occur in our studies, the magnitude of effect of volitional control observed in this research represents an inappropriately conservative estimate of the effect of free will. That is, the exercise of free will might actually reduce the magnitude of the volitional component observed.

Criticisms of a "Volitional" Interpretation

A second set of investigations probed the plausibility of the volitional account of the findings of the peanut-eating studies, as opposed to an efficient cause explanation of the data. Specifically, it was suggested by some that the differences in the amount of peanuts consumed on days when subjects were instructed to "eat" versus when they were told to "not eat" did *not* represent evidence for volitional control. Rather, critics saw these differences as confirming that peanut-eating behavior was under the control of the experimenter, because the subjects obeyed the experimenter's instructions. Or as one commentator put it, "The psychologist reader, steeped in behavioristic lingo, 'sees' in the experimental instruction to eat or not eat a *nonvolitional* control or manipulation directing behavior."

Two objections, which are really two ways of wording the same problem, were raised. These explanations suggest that the subject is either trying to be a good subject (Orne, 1962; Weber & Cook, 1972) or is trying to avoid being "socially sanctioned" for being an inconsistent, and therefore, a poor subject (Hayes, 1987).

First, we wish to address the contention that subjects are responding to demand characteristics, that is, that subjects had determined the experimenter's hypotheses and were behaving in such a manner as to confirm the hypotheses (and thus be good subjects). Note that the force of this position evaporates if one sees subjects volitionally choosing whether to adopt "good subject" or "bad subject" roles (See Howard, Youngs, & Siatczynski, 1988, Study 2). But, it also is worth mentioning that Howard and Conway (1986) note enormous differences among subjects, such that several subjects did not show any evidence of trying to play the role of "good subject."

Next, Hayes's (1987) reply focuses on research on "social standard setting," which speaks to our findings. Hayes believes subjects are able to behave in particular ways in experimental settings because they are "public" settings. But in "private" (read, nonexperimental) contexts, these subjects would be unlikely (perhaps even unable) to behave in the same manner. Although Hayes's research highlights an interesting factor in this domain, what remains unclear is exactly how important the public versus private nature of the data is in our studies.

Consider the following thought experiment. You set a jar of peanuts on your desk and flip a coin each day to determine whether it will be an "eat" or "try not to eat" day. Then you choose a colleague to tell which set of instructions you are entertaining, and you show him or her your daily results (public phase). Our data suggest that you might consume about six times the weight in peanuts on "eat" days as on the "try not to eat" days. Next, inform your colleague that your personal experiment has been completed, but then you continue the same set of procedures without telling anyone else (private phase). We would not be at all surprised if you found that the effect of volitional control remained unchanged in the private phase. Nor would we be concerned if, for example, your average weight consumed on "eat" days was only (for example) four times the amount consumed on "try not to eat days." The decline (from a sixfold to a fourfold advantage for "eat" days) represents an estimate of the impact of the private versus public nature of the experimental context.

When the present author conducted this experiment on himself, the advantage of "eat" days in the private phase was actually greater than the advantage of the "eat" days in the public phase. But I am completely confident that if Hayes conducted the same experiment upon himself, he would be perfectly able to show no effect of volition in the private phase—if he wanted to demonstrate such an effect. What would surprise me is if you (the typical reader) were unable to eat more peanuts on "eat" days than on "try not to eat" days in the private condition.

Of course, there is a simpler way of thinking of Hayes' public versus private challenge to our studies of volition. If one considers the issue to be one of external validity, the findings might be summarized as follows: In public settings (such as therapy), subjects are able to achieve their agentic goals more satisfactorily than in private contexts (such as self-improvement efforts; see Howard, DiGangi, & Johnson, 1988). Thus, if one is interested in public, psychological activities such as psychotherapy, the public studies we presented likely possess greater external validity than the private research alternative.

Empirical Investigations of the Obedience Hypothesis

Because of the possible plausibility of the conformity objection to the original volition studies, the following series of investigations explicitly attempt to test the volitional interpretation versus the "control via the experimenter's instructions" interpretation. In one study (Howard & Conway, 1986, Study 2), subjects sometimes received their daily "eat" or "not eat" instructions via a coin toss, and the experimenter was unaware as to what condition the subject was in for that day. Subjects still demonstrated an equally strong effect of volition. At other times, subjects simply chose and recorded whether that particular day would be an "eat" or a "not eat" day (and also did not let the experimenter know the condition they chose). In this condition, subjects also showed strong volitional control. Therefore, subjects showed volitional control in two different types of situations in which the experimenter not only did not give the "eat" and "not eat" instructions, but also the experimenter was not aware of whether subjects were in an "eat" or "not eat" condition on any day.

In yet another study, Howard, Youngs, and Siatzcynski (1988, Study 2) had subjects choose and record each day whether they would "follow instructions" or "do the opposite" of the instructions given (the meta-volitional factor). Subjects were then told by the experimenter either to "eat" or "not eat" peanuts that day. As expected, the volition by meta-volition interaction was significant and accounted for 65% of the within-subject variance. That is, subjects ate many more peanuts on "eat" days than on "not eat" days (135 g versus 10 g) when they had chosen to follow the instructions. However, when they decided to "do the opposite" they consumed far more peanuts in the "not eat" condition than in the "eat" condition (120 g versus 3 g). These studies serve to lessen the plausibility of the "they were compelled to obey the experimenter's instructions" efficient cause objection to the volitional interpretation of the above studies.

The intriguing aspect of the studies of volition reviewed above is that through the random assignment of conditions of volitional control (e.g., "eat–not eat," "binge–not binge," "initiate conversations–do not initiate conversations," "exercise–do not exercise") to time blocks, all possible explanations for mean differences between the two conditions, save two, are rendered implausible. The two possible explanations are (1) that these mean differences reflect the agents' power of self-determination (or behavioral freedom, or volition) in this particular instance and (2) that subjects were compelled to obey the experimenter's instructions and could not do otherwise (or, similarly, they had been so thoroughly socialized into a "good subject" role and therefore could not choose to disobey). Thus, the crux of the difference between these two explanations involves whether subjects obey their own directives or the directives of the experimenter. Howard (in press) addressed the above question of who causes the subject's behavior (the subject him/herself or the experimenter through the experimental instructions to the subject) by *collapsing the distinction between the subject and the experimenter.* Thus, the author served as both experimenter and subject for the study. Enormous volitional control of alcohol consumption was evident in this study. But if as both experimenter and subject he was merely conforming to the experimental instructions, then he was conforming to his own commands—but this is precisely the character of volition or self-determination!

Volition as a Factor in Therapeutic Interventions

We will now turn to applied and/or therapeutic investigations and applications of the above designs. Many psychologists have observed that self-determination is a critical factor in counseling and psychotherapy. Schultz (1977) concludes, after drawing from the collected theories of Gordon Allport, Carl Rogers, Erich Fromm, Abraham Maslow, Carl Jung, Viktor Frankl, and Fritz Perls, the following:

> Perhaps the only point on which they agree fully is that psychologically healthy persons are in conscious control of their lives. Healthy persons are capable of consciously, if not always rationally, directing their behavior and being in charge of their own destinies.

Thus, it has long been held by therapists that many of their clients' troubles stem from their inability to control either their environment or their response to the environment (Mahoney & Thoresen, 1974). The process of therapy often involves the process of reestablishing such control.

After having demonstrated a methodology for empirically assessing volitional effects in a series of studies on peanut eating, one reviewer wondered whether the concept of volition generalizes beyond the shell of a peanut. Indeed, my colleagues and I have considered the role of volition in domains that are of more interest to practicing psychologists. Many practitioners (e.g., clinical, counseling, industrial/organizational, and school psychologists) are often involved in behavior that their clients have difficulty controlling. At first blush it might appear that volition would not be a useful construct to aid these applied psychologists in their ministrations. For example, simply telling a schizophrenic to "stop hearing those voices," or instructing a depressed client to "feel less depressed," would seem to be singularly unhelpful. However, a slight modification of the procedures employed in the "peanuts studies" of volition has yielded some interesting findings in several clinically important domains. Rather than attempting to volitionally control the problem behavior directly, these studies adopt a strategy similar to that practiced by most experienced clinicians. Namely, subjects were encouraged to exercise volitional control *over the conditions that serve to maintain the problem behavior.*

We would like to suggest that the following studies be viewed as the beginning of a bridge-building process that will establish a direct link between research and the practice of therapy from a volitional perspective. The following sections highlight the issues addressed by volition research to date.

Frequency of Heterosexual Social Interaction

In one study (Howard & Conway, 1986, Study 3), college students who wished to increase the frequency of their heterosexual social (heterosocial) interactions were recruited for a study in which they were encouraged to exert volitional control over three factors related to heterosocial interactions: (a) the number of conversations initiated with members of the opposite sex, (b) the amount of time spent in places where social interactions frequently occur (e.g., dining hall, student center, parties), and (c) the frequency of positive self-statements students make about themselves and their social skills. Control of all three factors was structured in the same "try to . . . /try not to . . ." paradigm used in the peanut studies. It should be noted that, although few subjects in the previous "peanut studies" doubted their ability to control peanut consumption, the subjects in this study identified heterosexual social skills as an area of concern.

The data revealed that on the group analysis level, subjects were able to control all three of these conditions related to the number of heterosocial interactions, and that in so doing they were extremely effective in achieving their goal of having more (and more satisfying) heterosocial interactions.

Control of Snacking and Exercise

Lazarick et al. (1988, Study 1) considered the degree to which eating and exercise habits are under an individual's volitional control (for control of exercise, see also Howard, DiGangi, & Johnson, 1988). Subjects were divided into two groups, those who wanted to lose weight and those who did not particularly care to lose weight. After an initial baseline observations period, subjects were given a container of vegetables and one of four sets of instructions: (1) Snack on as many vegetables as you like but try not to exercise, (2) try not to snack on the vegetables and also try not to exercise, (3) snack on vegetables and exercise as much as you wish, and (4) try not to snack on the vegetables but exercise as much as you wish. The order of presentation of the four conditions was counterbalanced across subjects.

The results indicated that subjects could control both snacking on vegetables and exercising. Group differences were nonsignificant for the "desire to lose weight" variable. It should also be noted that the volitional control over snacking and exercising did not translate into weight loss as clearly as was hoped.

Control of Time Spent Researching Vocational Information

In another study reported by Lazarick et al. (1988, Study 2), the researchers considered how pursuing two types of information, vocational and personal, impacted on career indecisiveness. Subjects were given two information packets. The *vocational packet* contained information and exercises on the meaning of work, inventories pertaining to work values, talents, interests, and so forth. The *personal packet* covered topics like anxiety, goal setting, self-confidence, self-identity, and so forth. Subjects were then assigned to one of three conditions: (1) search for personal information, (2) search for vocational information, and (3) try not to search (purposely avoid attempting to discover about oneself or the world of work). The dependent measure was time spent engaged in the searching activity.

The study's results indicate that not only do subjects have a considerable amount of control over their behavior, but they were able to follow the "try not to search" instruction even though they reported dissatisfaction with this condition.

Control of Bingeing Behavior by Bulimics

Two studies (Lazarick et al., 1988, Study 3; Steibe & Howard, 1986) in this series have obtained strong evidence for the efficacy of a volitional treatment of binge eating and bulimia by using a "try not to binge" or "act normally" paradigm. The findings demonstrate that frequency of binge-eating episodes can be reduced and subjects can replace high-calorie foods with vegetables. Thus, binge eating is modifiable in the short run by efforts of will, but other issues remain unresolved making long-term control unlikely.

Control of Social Consumption of Alcohol

Finally, Howard (1986, 1988) and Howard, Curtin, and Johnson (1988) showed, with a slight variation in methodology, that social drinking is, indeed, under volitional control. In Howard, Curtin, and Johnson (1988, Study 1), subjects exerted control over their consumption of alcohol. The study had two phases. First, the subjects were to follow a baseline monitoring period. The second phase was the target hitting period in which they were to drink only as many glasses of alcohol as were indicated by the predetermined targets. Inaccuracy scores for the target-hitting period could then be compared to inaccuracy scores from the monitoring period. The latter were determined by comparing "normal" drinking patterns to a set of randomly selected targets.

To operationalize volition, then, subjects converted a significant message (FREE WILL), via Morse code, into target numbers. That is, a two-drink target represented a "dash," a one-drink target depicted a "dot," a zero-drink target represented a space between letters, and two consecutive zero-drink targets signaled a break between words. The subjects' volition was measured by how close they came to spelling "FREE WILL" and by how much they improved their accuracy from the baseline period to the target-hitting period. (It should also be noted that half of the subjects knew that they were spelling the words "FREE WILL" [the meaning condition], whereas the other half knew only the target numbers [less meaning condition].)

The results were striking. For example, those in the high-meaning condition achieved 100% accuracy and those in the low-meaning condition, although unable to spell the words precisely, improved their accuracy from baseline to intervention by up to 86%. Thus, these results invite another round in the controversy over whether the increased use of alcohol over time by nonaddicted individuals is best understood as a form of progressive physical addiction, a breakdown of volitional control, or both. Future research in this area with special populations and with the control of diverse environmental variables should prove both interesting and fruitful.

Finally, returning to the questions posed in this chapter's title, we can now respond that individuals are capable of controlling their actions to a large degree even when all possible nonagentic causes are methodologically controlled. In some instances (people who report drinking problems, participants' control over Internet use), the amount of control they exhibit is vanishingly small. In other instances (peanut eating, drinking alcohol by people who say alcohol is not a problem), the proportion of variance attributable to self-determination is enormous, and the amount of variance left for nonagentic causes is very small. Thus, human action appears to be partially self-determined and partially non-agentically caused. The instances of completely self-determined actions and totally nonagentically caused behaviors appear to be few and far between.

THE PLACE OF AGENCY IN PSYCHOLOGY'S FUTURE

Welcome to the 21st century. The psychologies of the 20th century often involved individual actions that helped to produce consequences *for that individual*. If I study hard for a test, my mark will improve, but my classmates' marks will be relatively unaffected by whether or not I study. Similarly, if my classmates exercise regularly, they will look better and be healthier, but I won't look better or be more healthy because they exercised. Individual actions (or lack thereof) led to consequences for that individual, but did not affect others. Environmental problems—which will assume far greater importance in the 21st century—are often different from the individually oriented "problems–solutions–consequences" nexes that psychologists encountered during the last century.

The amount of stress I (I) placed upon any ecosystem is given by the formula I = PAT, in which the stressors are the size of the human population (P) multiplied by the level of affluence (A), which is then multiplied by the environmental destructiveness of the technologies (T) used to supply that affluence (Ehrlich, 1991). Although individual-level behaviors (e.g., the size I choose for my family) always lead to environmental consequences, it is *not* my behaviors that mete out my consequences. Rather, it is what all members of the group do that determines the consequences that all of us experience. For example, it is relatively inconsequential whether I father 2 or 10 children. We will live in an overpopulated ecosystem based upon the actions of all humans in the ecosystem. Thus, if the average number of children born to fathers in an ecosystem is 2, the population's size will actually decline slightly, ceteris paribus (e.g., no net immigration inflow). Thus, what I do (e.g., 2 versus 10 children) is relatively unimportant compared to what all of us decide to do. The opposite is also true. If the average number of children in a family is 10, then we will all live in an overpopulated world regardless of my individual virtue or vice (e.g., 2 versus 10 children).

Because 20th-century problems usually linked individual actions to consequences borne by that same person, we could rely on a certain amount of self-interested motivation to help solve particular problems. But when all consequences are a function of group level actions, can (and will) individuals monitor their own actions and behave appropriately?

Economists have even specified the optimal strategy for rationally self-interested individuals, in our present world in which individual acts are virtually independent of group consequences, in the *free rider problem*. Specifically, I should enjoy my vices (e.g., have 10 children, take a helicopter to work) while encouraging all others to lead lives of virtue (e.g., have 2 or fewer children, slog to work via public transportation). In American society, such disciplines can be enforced economically—if they are enforced at all. That is, the price of owning a helicopter keeps most of us tied to the ground on our trip to work. Conversely, our present tax code grants much larger deductions (3 times as large) for families who have 10 children (12 exemptions) than for families with only 2 children (4 exemptions).

Obviously, the meaning of freely willed, environmentally charged acts in our present century is somewhat clouded. Are humans similarly motivated to act when consequences are meted out based upon group-level behaviors as they were when individual actions created consequences that were visited upon that actor? Is free will an endangered species of concept in a world dominated by vastly unequal economic opportunities? Or are humans simply incapable of self-determining their actions in an environmentally pressured world?

Psychologists who have studied the intricacies of environmentally important actions report that the problems encountered in fostering virtuous behaviors are great indeed (Howard, 1993, 1997, 2002, 2006; Winter, 1996). A frequent finding is that Americans are forced to choose environmentally appropriate acts in an economic environment that forces them to behave inappropriately. For example, as mentioned earlier, our tax code favors large families over appropriately sized families. Similarly, I must now pay 33% more each month if I choose to recycle part of my solid waste, than if I recycle nothing. Finally, who would pay several thousand dollars extra for a hybrid auto back when the price of gasoline was (artificially, through subsidies and tax allowances to the oil companies) kept well below $1.50/gallon?

In the world of environmentally important actions that will determine our individual and collective futures, it is appropriate to ask, "Exactly whose will is being served by our present tax code, waste removal systems, energy price arrangements, and the like?" And exactly how free are Americans to counteract these economic forces and create an environmental world suitable for our children? Finally, why is our present political system hell-bent on preserving 20-century energy systems (i.e., transportation by burning oil-based fuels, electricity produced by burning coal, etc.) for the overpopulated world of tomorrow? And the most important question of all, what prevents us from changing this mad dash toward environmental insanity?

The research reported earlier on self-determination by individuals demonstrated that people could alter the conditions that influence one's behavior (e.g., by spending more time in social places, a person will have more heterosocial interactions). This "indirect method" of determining my actions appears to be an important strategy that people sometimes employ to behave volitionally. But when one considers the conditions related to important environmental actions (e.g., family size, recycling, type of transportation I use) one can see that I have little or no ability to control the relevant conditions (e.g., the tax code, the pricing system for waste removal, whether auto manufactures produce hybrid autos). Does this mean we have less power to self-determine our actions? Do we therefore have little ability to change our actions in domains important to the environment?

The problems that individuals faced in the last century were difficult enough to tackle. Perhaps the next century's problems will be more difficult still. Finally, it is clear that it will take "political will" to change conditions like the tax code, waste payment systems, and the mix of high- versus low-mileage vehicles that companies offer. I trust that humans will self-determine their actions in the 21st century, just as they did in earlier times. However, I fear that in some domains (e.g., environmental issues) gaining the ability to self-determine our actions might prove to be uncommonly difficult.

REFERENCES

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Ehrlich, P. E. (1991). *Healing the planet*. Reading, MA: Addison Wesley.

Hansen, N. R. (1958). *Patterns of discovery.* Cambridge, England: Cambridge University Press.

Hayes, S. C. (1987). Contextual determinants of "volitional action": A reply to Howard and Conway. *American Psychologist, 42*, 1029–1030.

Howard, G. S. (1986). *Dare we develop a human science?* Notre Dame, IN: Academic Publications.

Howard, G. S. (1988). Science, values and teleological explanations of human action. *Counseling and Values,32*(2), 93–110.

Howard, G. S. (1989). *A tale of two stories: Excursions into a narrative approach to psychology.* Notre Dame, IN: Academic Publications.

Howard, G. S. (1993). *Ecocounseling psychology. The Counseling Psychologist, 21*, 549–616 (whole issue).

Howard, G. S. (1994). Some varieties of free will worth practicing. *Journal of Theoretical and Philosophical Psychology, 14*, 50–61.

Howard, G. S. (1997). *Ecological psychology: Creating a more earth-friendly human nature.* Notre Dame, IN: University of Notre Dame Press.

Howard, G. S. (2002). *How should I live my life? Psychology, environmental science, and moral traditions.* New York: Rowman & Littlefield.

Howard, G. S. (2006). *Stan Ovshinsky and the hydrogen economy: Creating a better world*. Notre Dame, IN: Academic Publications.

Howard, G. S., & Conway, C. G. (1986). Can there be an empirical science of volitional action? *American Psychologist, 41*, 1241–1251.

Howard, G. S., Curtin, T. D., & Johnson, A. J. (1988). *The hardening of a "soft" science.* Invited Address: Mathematical and Statistical Models of Behavior Track; Science Weekend, APA Convention, Atlanta.

Howard, G. S., DiGangi, M. L., & Johnson, A. (1988). Life, science, and the role of therapy in the pursuit of happiness. *Professional Psychology: Research and Practice, 19*, 191–198.

Howard, G. S., Youngs, W. H., & Siatczynski, A. M. (1988). Reforming methodology in psychological research. *Journal of Mind and Behavior, 10*, 393–412.

Kuhn, T. S. (1962). *The study of scientific revolutions.* Chicago: University of Chicago Press.

Kuhn, T. S. (1977). *The essential tension.* Chicago: University of Chicago Press.

Lazarick, D. L., Fishbein, S. S., Loiello, M. J., & Howard, G. S. (1988). Practical investigations of volition. *Journal of Counseling Psychology, 35*, 15–26.

Mahoney, M. J., & Thoresen, C. E. (1974). *Self-control: Power to the person*. Monterey, CA: Brooks/Cole.

Orne, M. T. (1962). On the social psychology of the psychological experiment. *American Psychologist, 17*, 776–783.

Schultz, (1977). *Growth psychology: Models of the healthy personality.* New York: D. Van Nostrand.

Skinner, B. F. (1971). *Beyond freedom and dignity.* New York: Knopf.

Steibe, S. C., & Howard, G. S. (1986). The volitional treatment of bulimia. *The Counseling Psychologist, 14*, 85–94.

Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research. *Psychological Bulletin, 77*, 273–295.

Winter, D. D. (1996). *Ecological psychology: Healing the split between planet and self.* New York: HarperCollins.

14 Free Will as a Proportion of Variance

William R. Miller
David J. Atencio

The historic debate regarding human volition has often been cast as a dichotomous choice between free will and determinism. At one extreme of this polarity, humans are entirely self-directed, choosing their own thoughts and behavior unfettered by nature or nurture. At the other extreme, choice and self-determination are regarded to be illusory, and all human action is the cross-product of biology and experience, with a random error term.

Neither extreme is viable. Psychological science amply demonstrates the influence of experience and contextual factors on behavior. Behavior genetic research estimates the proportion of variance in human traits attributable to our chromosomal heritage. Clearly it is infeasible to argue that we are uninfluenced by our genes, environment, and experiential history.

Being influenced, however, is not the same as being determined. The recognition of human choice is deeply embedded in natural language, subjective experience, ethics, morality, and law. The ability to choose between good and evil, between right and wrong paths, is fundamental in all major world religions (Miller & Delaney, 2005). Criminal law is also predicated on the assumption that people can and do choose their actions from among alternatives. Special mens rea provisions increase or decrease personal culpability on the basis of choice, with more severe penalties for behavior that is judged to be intentional and planful (e.g., premeditated murder). There are certain legal conditions of heightened accountability for choice (e.g., lying under oath). Conversely, in English and American law a person found to be incapable of knowing right from wrong is regarded as not guilty (or in Scottish legal tradition, guilty but insane) and therefore not deserving of punishment. The law also recognizes degrees of diminished capacity for planful action. Coercion ("I had no choice") is taken into account in human discourse. Words uttered or signed with a gun to the head are not so binding or blameworthy as language spoken in the absence of constraint. Witnesses to the signature on a will attest that the signer was of sound mind and under no apparent coercion. Volitional choice is a fundamental foundation of accountability and a civil society.

In essence, humans author their lives but never start with a blank page (Pinker, 2002). If volition or will influences human behavior (as we here assert that it does), it is but *one* determinant, albeit one to which society ascribes particular importance. In other words, volitional choice accounts for some *proportion* of the variance in behavior. Hereditary and environmental factors also shape behavior and share variance with volition. Human behavior is influenced by and also alters context, a cycle described by Bandura (1986) as reciprocal determinism. This same reciprocity operates at a social level as well, with culture both shaping and shaped by human evolution (Baumeister, 2005a).

VOLITIONALITY

Within this perspective of will as one determinant of human action, it is conceivable to estimate a proportion of variance in behavior for which volition accounts (Howard & Conway, 1986; Miller & Brown, 1991), much as behavior geneticists project the heritability of behavior. This might be called the *volitionality* of behavior, the extent to which a behavior is subject to willful control. Volitionality surely varies across behaviors, some of which are much more readily self-regulated than others. One index here is the extent to which a behavior is subject to instructional control. Most people can readily comply when asked to close their eyes, raise a hand, or sample from a bowl of chocolates on the left rather than from the bowl on the right. Fewer can comply when asked to decrease their heart rate, do a handstand, or not think about raccoons. The controllability of behavior in response to instruction also seems easier to ascertain than whether an individual behavior was chosen (volitional) on a particular occasion.

The extent of volitional control also varies across individuals (Snow, Corno, & Jackson, 1996). Some people evidence greater self-control across a range of behaviors, in contrast to those who show more general impulsivity (McRae & Costa, 2003). By practicing self-regulation in the face of temptation, individuals can gradually build up greater volitional "muscle" that can be applied in a wide variety of situations (Baumeister, 2005b; Baumeister, Heatherton, & Tice, 1994). Beyond generalized self-regulation, the controllability of specific behaviors also varies across individuals. A classic example is alcohol consumption, which most adults have little difficulty self-regulating, at least within a culture in which moderate alcohol use is normative. Other individuals, however, experience subjective loss of control when using alcohol (or some other psychoactive drug), making self-regulation more difficult (albeit not impossible) after an initial dose (Bechara, 2005; Fingarette, 1988). *Addiction* is a popular term to describe such encapsulated areas of diminished volitional control over behavior.

Finally, the controllability of a specific behavior can vary within the same individual across time. The development of addiction involves a gradual erosion of volitional control. The proportion of variance accounted for by personal choice diminishes over time, as the behavior becomes more frequent and predictable (Bechara, 2005). Failed efforts at self-control are symptomatic of substance dependence (American Psychiatric Association, 1994). Volitional control is never entirely absent in addiction, but automaticity increases as biological and environmental determinants gain hegemony (Miller & Carroll, 2006).

TOWARD A PSYCHOLOGY OF VOLITION

Despite or perhaps because of the historic debate regarding free will versus determinism, a coherent psychology of volition is still nascent. Although concepts of agency and will have been the subject of philosophical debates since antiquity (Baumeister, 1987; Broughton, 1986), human agency has proved challenging to study directly in behavioral science (Howard, 1986; Wertsch, Tulviste, & Hagstrom, 1993). It is striking that a concept so universal in human experience and so fundamental to social discourse should be relatively ignored within psychology, the mainstream science of human nature (Miller & Delaney, 2005).

Volitional Action

What constitutes volitional control over behavior? There is no clear scientific consensus as to what defines intentionality when observing overt behaviors (Lutkenhaus & Bullock, 1991). Social judgment, however, routinely includes evaluation of the extent to which an action was volitional, and language contains a rich vocabulary to describe this domain: intentional, on purpose, calculated, conscious, manipulative, deliberate, willful, and so on. This judgment is formally codified in overlapping legal concepts of mens rea such as specific and general intent, malice aforethought, diminished capacity, voluntariness, and insanity. The logic of volition in social judgment, a shared social "theory of mind" is thus one source to inform a psychology of volition. Within a culture, at least, people judge intentionality according to specifiable rules, and do so with a high degree of agreement (Malle & Nelson, 2003).

To be sure, there is a clear human tendency to overestimate the contribution of volition to behavior. First-person linguistic accounts provide explanations of actions that are, in essence, retrospective causal hypotheses (Harré & Secord, 1972). People manufacture plausible explanations for behavior that is demonstrably controlled by factors of which they are not consciously aware (Bargh & Chartrand, 1999; Bargh & Ferguson, 2000). People also tend to explain their own situationally predictable behavior in volitional terms, whereas to observers the influence of environment is more salient (Jones & Nisbett, 1972). Human beings do often behave in ways that are consistent with demand characteristics of the situation, and may be unaware of these determinants of their behavior (Ross & Nisbett, 1991). This was a primary source of the humor in the television program *Candid Camera*. That such misattribution occurs *sometimes* is by no means proof for the determinist perspective that *all* human behavior is driven by nature and nurture. Indeed, psychology is relatively late among scholarly disciplines in recognizing logical positivism as a failed model of human nature, precisely because it does not take into account that which is uniquely human.

What is needed, then, is a psychology of volition that incorporates (rather than competes with) the influences of hereditary and environmental factors. The choices that people make are surely shaped and constrained by nature and nurture. What we assert is that the static facts of heredity and of past and current environment are only partial determinants of behavior, and any picture of human behavior that is limited to these elements is necessarily incomplete. A portion of the variance in human behavior, often a substantial portion, is linked to volition—to choices (decisions, plans, commitments, etc.) that the person has made. From among available courses of action, the person chose this particular option but could have done otherwise. As noted above, this is a foundational assumption in much of human discourse including business, law, and religion. The pragmatic relevance of choice is apparent in the vast sums spent to influence it in advertising and politics.

We suggest that volitional action is constituted of two necessary components: an intention, and an enactment of that intention. Neither unconsummated intention nor unintended action is sufficient. Considerable progress has been made in understanding both how cognitive processes govern the formulation of intentions that lead to specific behaviors (Ajzen, 1985; Ajzen & Fishbein, 1980) and the cognitive control and maintenance of actions once enacted (Kuhl & Beckmann, 1985).

Intention

Intention is a necessary but not sufficient component of volition (Zhu, 2004). An act cannot be volitional if it was unintentional, if the person did not "mean" to do it. Juries are regularly asked to decide whether an action was deliberate (e.g., intentional, premeditated), a judgment with implications for the defendant's guilt or degree of culpability. Homicide, for example, is differentiated in American law into three levels of severity: unintentional (third degree), intentional (second degree), or deliberate and premeditated (first degree). For certain other crimes, guilt depends upon the person's ability to have formed a specific intent to commit the act, and for still others, culpability can be reduced by the demonstration of diminished general capacity for intentional action. As amply demonstrated in research on the influential theory of planned behavior (Ajzen, 1991), intentions are a reasonably good predictor of future behavior. Stated intentions (Gollwitzer, 1999; Gollwitzer & Schaal, 1998; Rise, Thompson, & Verplanken, 2003) and commitment (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003) are reliable predictors of subsequent enactment.

Intention begins with desire. No action is likely to be judged volitional unless it is linked to a goal, something that the person hoped for or wanted. Volitional behavior is directed toward a desired outcome; in legal terms, a motive.

A desired outcome is disembodied, however, without a means to achieve it. A second component in intention is the belief that there are effective means to achieve the goal, that there is behavior that is at least reasonably likely to have the desired result. This is what Bandura termed *general efficacy* (Bandura, 1982). Given the existence of potentially effective behavior, an individual may or may not perceive personal ability to carry out the behavior and succeed—that which Bandura called *self-efficacy.* If a person desires an outcome, knows that there are behaviors to achieve it, but believes that these are beyond his or her personal capability, no intention is formed. (Here we disagree with Malle and Nelson's [2003, p. 568] statement "If skill plays a role, it can only be a condition of intentionally performing an action, not a condition of forming an intention.")

With these pieces in place—a goal or motive, a potentially effective action, and personal efficacy to complete the action—the preliminary ingredients of intention are available. What remains is the decision to proceed, the formation of intent. This developmental stream is illustrated in the figure on page 280.

So how is intention normally evaluated in social judgment? One influential factor is a person's stated desire or intention to commit the act (or achieve a consequence). A prior statement such as "I'm going to kill him" is material evidence in judging volitionality, as is a post-hoc description of purposeful motivation. Lacking such a direct statement, intention is also inferred from apparently planful antecedent steps leading up to the action, suggesting that the behavior was consciously directed toward a goal. The person is likewise presumed to have freely chosen this course of action from among alternatives (i.e., could have acted otherwise).

Components of Volitional Behavior	*Corresponding* **Mens Rea** *Concepts*
Desire (Goal)	***Motive***
⇓	
Effective Means (General Efficacy)	***Aforethought***
⇓	
Personal Ability (Self-Efficacy)	***Deliberately, Purposely***
⇓	
Intention (Commitment)	***Willful, Planful, Intent***
⇓	
Enactment with Awareness	***Knowingly, Voluntarily***
⇓	
Consequence	

Language is a uniquely human influence on behavior, and is the domain within which intention resides (Harré & Secord, 1972). People not only act, but talk to themselves and others about their actions. Verbal instructions obviously can be a sufficient stimulus to produce specific behavior. Internal private speech similarly shapes behavior (Diaz, Winsler, Atencio, & Harbors, 1992; Meichenbaum, 1977; Meichenbaum & Goodman, 1971). If prediction and control are the substance of science, then scientific demonstration of volitional control is relatively straightforward: Individuals can predict or self-control their own behavior, albeit imperfectly (Howard, 1986). This permits estimation of the proportion of variance in behavior that could be accounted for by volition. A series of illustrative experiments demonstrated volitional control of peanut eating by experimenter instructions, by participant self-declared intention, and in relation to a coin toss with results unknown to the experimenter (Howard & Conway, 1986). Other studies have shown instructional control of exercise, eating vegetables, and seeking vocational information (Lazarick, Fishbein, Loriello, & Howard, 1988). Prospectively, the voicing of intention and commitment increases the probability of subsequent behavior (Amrhein et al., 2003; Gollwitzer, 1999), though of course the linkage is less than complete. All of these phenomena bespeak planful, goal-directed action.

There is a meaningful (albeit sometimes moot) distinction here between volitional action and intended consequences. A person may have decided to take a certain action (e.g., to drive a vehicle while intoxicated) without intending the particular consequence that ensued (e.g., injury and death). In social judgment, people are often held responsible for unintended consequences, although in some circumstances the perceived degree of fault is diminished (e.g., in jury assessment of punitive damages).

Enactment

Volitional action involves not only intent but also behavioral enactment of that intention. Volitionality requires forming an intention or decision and then initiating and sustaining action in accord with it (Perugini & Bagozzi, 2004; Searle, 2001). Time-framed intention is the defining difference between contemplation and preparation stages in the transtheoretical model of change, whereas it is enactment that characterizes the subsequent action stage (Prochaska, 1994; Prochaska & DiClemente, 1984). Legal sanctions typically apply to what an individual actually does, and not to mere intentions. (One departure from this is recent U.S. responses to terrorism, including incarceration and invasion based on inferred intent to do harm.) Volition thus involves both an antecedent intention and goal-directed behavior.

Not all goal-oriented behavior is volitional. Rapidly withdrawing one's hand when inadvertently touching a hot surface is an adaptive reflex, hardwired to prevent injury and promote survival. It occurs quickly "without thinking," and thereby most would not judge it to be an act of will. In contrast, holding one's hand in flame has the appearance of volitional action, contravening as it does a high-probability response of self-protection. Taking the road less traveled enhances the perception of intentionality. One possibility, then, is to regard volition as a residual error term, the amount of variance that is unexplained by known determinants. This is unsatisfactory, however, because people also can and often do choose the predictable path (which is what defines it as predictable). To go with the flow, to do what is consistent with contextual demands such as peer pressure, does not obviate volitional choice and responsibility. Thus, there is shared variance between environmental and volitional influences on behavior. Furthermore, trait volitionality is itself surely influenced by genetics: Some people are constitutionally more capable of overriding salient situational cues, and volitional unpredictability may have evolutionary advantages (G. F. Miller, 1997). Traits such as impulsivity, empathy, self-monitoring, and conscientiousness are genetically linked phenotypes likely to influence an individual's exercise of volition. It is unsatisfactory, therefore, to relegate volition to leftover variance, precisely because volitionality overlaps and interacts with both hereditary and environmental factors in shaping behavior.

Variability

Volitionality changes for the same individual and behavior across time and contexts. The concept of temptation bespeaks environmental conditions under which volitional control is more difficult. In the language of relapse prevention (Marlatt & Donovan, 2005), these are high-risk situations in which there is greater danger of self-regulation failure (Baumeister et al., 1994).

An individual's capacity for volitional control of a particular behavior can also vary over time. Hamlet advised his mother, "Refrain to-night, and that shall lend a kind of easiness to the next abstinence: the next more easy; for use almost can change the stamp of nature" (*Hamlet*, Act 3, Scene 4). The practice of abstinence itself diminishes vulnerability to temptation. The future stability of abstinence from alcohol, for example, increases with the duration of prior abstinence (Maisto, Clifford, Stout, & Davis, 2006). Following a span of abstinence, however, a single exception can trigger a rule-violation effect (e.g., "Now I'm off my diet") that compromises restraint (Marlatt & Donovan, 2005). Some behaviors become easier to self-regulate with practice; for example, where one hits a tennis ball when returning a serve. Other responses such as drug taking more quickly come under the control of biological and contextual factors, so that volitionality diminishes with practice.

Developing Self-Regulation and the Capacity for Volitional Action

Relevant in understanding volition is the developmental literature on self-regulation. Clearly human adaptive abilities develop through complex transactions in the physical and social environment (Ford, 1987; Gottlieb, 1992; Riegel, 1976). Self-regulation in all domains (e.g., behavioral, cognitive, emotional, conative, social) is a capacity that emerges in transactions between the developing child and the people, objects, and activities around them (Bronfenbrenner & Morris, 1998). This capacity to plan, organize, direct, flexibly adapt, and orchestrate one's thoughts, feelings, and actions in dynamic environments involves effortful control both to inhibit and to enact certain responses in pursuit of a goal (Rothbart & Bates, 1998).

As with other psychological phenomena, an understanding of self-regulation can be aided by examining the developmental sequence with which it emerges (Kopp, 1982, 1987, 1991). Neonates possess innate tendencies for homeostasis that involve attending, orienting, and responding to the environment. Young children develop the ability to comply with caregiver directives to inhibit or enact specific behavior. At first, such compliance depends upon the immediate presence of the caregiver, but with the development of language children gradually acquire the capacity to internalize rules and adhere to them even in the absence of the caregiver (Diaz et al., 1992). (A child approaches a hot stove, verbalizes "Hot, don't touch!" and pulls back.) As parents well know, older children move beyond compliance with externally directed rules, and become capable of making choices, monitoring and directing their behavior toward their own goals. In essence, volition emerges from being socialized by others, adapting to more complex environments, and gradually taking over the regulatory role as caregivers withdraw their control. Self-regulatory capacities become not only organized into a functional system but also more complex as the child moves from reliance on external structure to the ability to self-formulate goals and plans of action. The human capacity for self-determination is thus one that emerges over the course of development (Deci & Ryan, 1985; James, 1899; Nuttin, 1987). Various parenting styles can facilitate or hinder this developmental process, affording children differing opportunities to exercise behavioral, social, cognitive, and emotional self-organization (Aasor, Roth, & Deci, 2004; Baumrind, 1966; Diaz, Neal, & Vachio, 1991; Grolnick & Ryan, 1989; Kopp, 1987).

ENGAGING VOLITION

A variety of methods have been developed for bolstering willful self-regulation of behavior; in essence, for increasing the proportion of variance that is under volitional control. Our discussion here is divided into two sections—instigation to change and self-regulation—corresponding to the two above-described components of volition: intention and enactment.

Instigation to Change

Building on earlier work by Kanfer (1970), Miller and Brown (1991) hypothesized a seven-step homeostatic cycle of self-regulation. Incoming information about the person's current status (1) is compared with internal standards (2). So long as status is within acceptable limits, no change is required. A significant discrepancy between status and standard, however, constitutes a state that they termed *instigation to change* (3), which triggers a scanning for possible alternative courses of action (4). The detected options are evaluated, and a course of action is chosen (5) and implemented (6). Subsequent monitoring of changes in the degree of discrepancy (7) yields new information, thus returning to the top of the cycle (1).

Steps 1–5 of this model describe the formation of intention, the first of two components of volitional action. Status quo is the default. Once established, many behavioral routines go on automatic pilot, requiring little or no intentional control to be maintained. Instigation to change occurs when the person perceives a significant discrepancy between how things are and how they ought to be. This does not in itself constitute an intention, but it does activate a search for options that might reduce the perceived discrepancy between status and goal. If no possible, acceptable, and potentially effective course of action is identified, status quo remains. If, on the other hand, the evaluation of options yields one or more potentially effective courses of action that the person can (Bandura, 1982) and is willing to do, then an intention may be formed.

This conceptual model of self-regulation can inform methods for instigating change, which essentially aim to create a perceived discrepancy between status quo and desired goal (Miller & Brown, 1991). One such strategy is to provide informational input to be compared with internal standards. Credible feedback of one's current state (e.g., stepping on a bathroom scale) is an important source of motivation (DiClemente, Marinilli, Singh, & Bellino, 2001; Locke & Latham, 1990). Such feedback may be most useful in situations in which the person does not currently perceive any need for change, and when it provides information that is not already apparent to the individual (e.g., blood pressure reading or cholesterol level). The "drinker's checkup" was devised precisely for this purpose, offering people information about their volume of alcohol use relative to population norms, levels of problems and dependence, and the current effect of drinking on their liver and neuropsychological functioning (Miller, Benefield, & Tonigan, 1993). When compared with two established forms of psychotherapy, this brief checkup yielded similar reductions in drinking over 1 to 3 years of follow-up (Project MATCH Research Group, 1997, 1998).

A second possible strategy for instigating change is to alter the standard against which the current status is compared. For example, medical standards for healthy blood pressure and cholesterol levels have changed over the years, so that levels once regarded to be normal are now regarded as cause for concern. Standards for what constitutes normal drinking tend to be more liberal among heavy drinkers, and people more generally tend to overestimate the frequency of deviant behaviors in which they engage—a phenomenon known as the false consensus effect (Agostinelli & Miller, 1994). "Norm-correction" strategies have been tested, providing actual behavioral norms, and have been effectively combined with feedback of personal status relative to norms (Agostinelli, Brown, & Miller, 1995). The combination of creating a perceived need and a solution to decrease discrepancy has also been common in advertising and sales, in which case the intended instigation for change is selection of a particular product.

Instigation can also occur without new information, through internal processes of self reevaluation (DiClemente, 2003). Even without significant change in external reality, people may decide that a change is needed (Premack, 1972). Motivational interviewing (MI; Miller & Rollnick, 2002) was developed specifically to help resolve ambivalence and promote instigation for change. Within a supportive and empathic counseling style, the person is encouraged to voice and explore his or her own motivations for change. MI has been found to promote change across a wide range of health behaviors (Hettema, Steele, & Miller, 2005). Psycholinguistic analyses of MI sessions suggest a sequence of reevaluation, whereby individuals first explore self-motivational themes including desire, ability, reasons, and need for change. This in turn leads to increasing strength of commitment or intention language, which predicts behavior change (Amrhein et al., 2003). Desire, ability, reasons, and need may not in themselves be sufficient to instigate change, but appear to be part of the process that ultimately leads to instigation.

Strengthening Self-Regulation

Clearly intention is not enough. After a person chooses and commits to a course of action, it remains to carry it out (Ach, 1910). Implementation offers its own challenges, particularly when the change requires not a one-time act but sustaining new behavior over time. Volitional processes around enacting, maintaining, and executing intentions differ from the antecedent motivational processes that form desires, goals, and intentions (Corno, 1993; Kuhl, 1985).

The concept of behavioral self-control arose in the late 1960s and 1970s, and exemplifies the interface of volitional and contextual influence (Kanfer, 1970; Mahoney & Thoresen, 1974; Thoresen & Mahoney, 1974). Within this perspective, people learn and apply behavioral science principles to help them carry out intended behavior change. Common strategies include self-monitoring, avoiding or altering sources of temptation, arranging reminder cues to prompt behavior change, interrupting habitual patterns of action, and rewarding new behavior (e.g., Miller & Muñoz, 2005). There are also strategies to "decide in advance" by placing planned restrictions on behavior, such as carrying a limited amount of money and no credit cards into a casino, or taking the medication disulfiram that causes illness if the person drinks alcohol within the next few days. In essence, such people are changing their environment in systematic ways in order to promote their own enactment of a volitional behavior change.

Promoting Integrity With Values

The above discussion has focused primarily on intention and implementation of a particular act. Daily life, however, involves simultaneous engagement with multiple short- and long-term goals that necessarily compete with each other for time and resources. Rokeach (1973) conceived of personal life goals as hierarchical, with a relatively small number of core values, a somewhat larger subservient set of instrumental values, and more peripheral supporting structures of attitudes and behavior patterns. In essence, one's core and instrumental values serve as guiding rules for living.

The larger virtues of integrity and self-control can be understood as the ability to conform one's behavior to be consistent with and promote one's central values. This can also involve the selection of which goals to prioritize, and in particular forgoing short-term goals of immediate gratification that interfere or compete with the actualization of longer-term values. The psychologist O. Hobart Mowrer, best known for his contributions to learning theory (Mowrer, 1983), in later life turned his attention to developing an *integrity therapy* designed to help people conform their lives with greater consistency to their central goals (Drakeford, 1967; Lander & Nelson, 2005; Mowrer, Vattano, et al., 1974; Mowrer & Vattano, 1976).

Integrity often involves willful adherence to values rather than to the demands, temptations, and distractions of daily life (Frankl, 1963). Fidelity to core values can at times require conscious, volitional overriding of other influences on behavior—in other words, behaving in rule-governed ways that would not be expected from the environmental conditions at hand, or even from prior behavior patterns (Frankl, 1969). Such exercise of volitionality in service to larger values is uniquely human.

An illustrative example is human recovery from addiction. One needs nothing more than animal learning principles and neurobiological models to understand how people can fall into substance dependence (Logan, 1993). Addictive behaviors involve direct (primary) and indirect (secondary) positive reinforcement for use, and sometimes negative reinforcement for use as well (avoiding aversive states). Add to this neuroadaptation that occurs with chronic substance use (Koob, 2005), and the incentives for continuation are formidable. Laboratory animals succumb to precisely these influences. There is, however, no animal model for sustained abstinence as is widely observed in treatment, Alcoholics Anonymous, and natural recovery. Such recovery involves the human frontal cortex, and volitional override of well-established and strongly reinforced behavior patterns.

COLLECTIVE WILL

Thus far our discussion has focused on free will in the behavior of individuals. Because individuals are embedded in social systems, it is also possible to consider the collective volition of groups. All persons travel unique life courses that intersect with others in complex, reciprocal, and often recurring ways (Taylor, 1991). To what extent can a family, community, or nation exercise collective will to override the expected product of nature and nurture, or to break a repetitive destructive cycle?

It is a natural step from individual agency to collective will. The very knowledge of self as well as one's sense of personal agency develop through social interaction and intersubjectivity (Bronfenbrenner & Ceci, 1994; Mead, 1934; Trevarthen, 1993). Like individuals, social systems are dynamic, with collective goals and the potential to exert influence on others (Bandura, 1986). Intention and enactment, the two components of volitionality for individuals, can also be applied to understand acts of collective will.

Collective Intention

Volition begins with an intention, which in turn is rooted in a goal, a perceived discrepancy between how things are and how they ought to be. Like individuals, groups have an identity and values, a collective sense of self that differentiates them from other groups (Spears, Jetten, & Scheepers, 2002). An important aspect of collective identity is shared goals; for example, the mission statement of an organization or the platform of a political party. Collective values may not be stated so formally, but can be inferred (as with individuals) from a group's actions. In fact, implicit values reflected in a group's actions may or may not conform to its formally stated values (Kuhl, 1985). A collective goal may also emerge spontaneously from shared circumstances, such as a group's desire to escape from a theater in response to a fire alarm.

Groups often form around shared goals. For example, communities of practice (Lave & Wenger, 1991) emerge when people join together because of common interest around some problem or subject and collaborate over extended periods of time to share knowledge, develop ideas, problem solve, and build innovations. Through the actions of a community of practice, participants engage in mutually shared activities in a collaborative manner that is determined by coordinated responsibilities and ongoing adjustments (Rogoff, Turkanis, & Bartlett, 2001; Wenger, 1998).

As discussed above, self-efficacy is a necessary component for a goal to become an intention. A perceived discrepancy between how things are and how they ought to be is unlikely to trigger action unless change is perceived as possible. Bandura (1997) defined collective efficacy as "a group's shared belief in its conjoint capabilities to organize and execute the courses of action required to produce given levels of attainments" (p. 477). Perceived collective efficacy is thus a central influence on what groups can and will do. Collective efficacy is comprised in part by the aggregate of group members' individual perceptions of their own individual capabilities to perform particular roles and functions necessary to achieve a goal. Also relevant is the aggregate of group members' individual perceptions of the group's capabilities as a whole (Bandura, 1995, 1997). Perceived collective efficacy influences what people choose to do as a group, the collective effort that is exerted in pursuit of shared goals, and the degree to which a group persists in face of challenges and even failure to produce results (Bandura, 1986).

These, then, are two necessary (but not necessarily sufficient) elements of a collective intention: a shared goal and perceived collective efficacy of the group to accomplish it. What moves the group to an instigation to change, a collective intent, and commitment to act? Clearly, collective instigation is often aided by the presence of a charismatic leader such as César Chávez, Martin Luther King, or Nelson Mandela. Sometimes the fuse is lit by a public event that brings a discrepancy into sharp relief and crystallizes discontent (Baumeister, 1994). Whatever the particular catalyst, goal-discrepancy and perceived efficacy coalesce to form a collective intention, a commitment to action.

Collective Enactment

Deciding to do something is not the same as doing it. At a collective level, the enactment component of will involves the extent to which the group is able to focus its intention and function together in pursuit of a shared goal. This depends in part on cohesiveness and the extent to which individual members identify with the group. Sports teams illustrate this cohesive identification process. Early in training, there is a focus on working together as a team, with individuals establishing their roles on the team. Ideally a new collective identity emerges, that of the individual-in-relation (i.e., myself as a member of the team).

Typically enactment requires continued effort over time, so that collective will requires not only initial resolution, but resolute persistence. Action is taken in pursuit of the goal, with resulting feedback as to whether the goal is closer to attainment. A team responds to the experience of each win or loss, possibly with new strategies, adjustments in practice, or reassignments of individual roles.

Collective will can be compromised when personal goals conflict with shared goals, and individual members disengage from or contravene the group's efforts and values (Bandura, 1990, 1991). Collective will, therefore, is also promoted by a group's ability to inspire and maintain its members' commitment to shared goals and efforts.

SUMMARY

Human behavior is directed not only by biological heritage and the current and past environments, but also by the characteristically human capacity for willful self-determination. This characteristic of human choice is assumed in morality, law, religion, business, and everyday social discourse. The capacity for self-regulation develops in childhood, first in the capacity to comply with instructions from others, and then in development of and adherence to personal goal-directed plans. Volitional action requires both the formation of an intention and enactment of planned action to fulfill the intent. Volition is but one influence, interacting with nature and nurture, and thus accounting for only a proportion of the variance in human behavior. Volitionality varies across individuals, across behaviors, and for the same behavior within the same individual across time and contexts. There are surely constraints on the proportion of behavior that can be self-regulated, and people tend to overestimate the degree to which their own actions are volitional. Freedom most always exists within limits. Will, in this sense, is not totally free. Instead it works in concert with, and sometimes overrides the expected effects of nature and nurture. This enduring tension in human experience is reflected in Reinhold Niebuhr's well-known serenity prayer, so widely used within Alcoholics Anonymous:

> God grant me the serenity to accept the things I cannot change,
> The courage to change the things I can,
> And the wisdom to know the one from the other.

Psychology needs to develop and integrate an understanding of volition as one significant determinant of individual and group behavior. Failing to do so, psychology defaults on its historic role as the science of the *psyche* and not merely of behavior, overlooking that which is widely understood to be a unique and defining attribute of human nature.

REFERENCES

Aasor, A., Roth, G., & Deci, E. L. (2004). The emotional costs of parents' conditional regard: A self-determination theory analysis. *Journal of Personality, 72*, 47–88.

Ach, N. (1910). *On volition.* Leipzig: Verlag von Quelle & Meyer.

Agostinelli, G., Brown, J. M., & Miller, W. R. (1995). Effects of normative feedback on consumption among heavy drinking college students. *Journal of Drug Education, 25*, 31–40.

Agostinelli, G., & Miller, W. R. (1994). Drinking and thinking: How does personal drinking affect judgments of prevalence and riskiness? *Journal of Studies on Alcohol, 55*, 327–37.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–39). Berlin: Springer-Verlag.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211.

Ajzen, I., & Fishbein, N. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice-Hall.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington, DC: American Psychiatric Association.

Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology, 71*, 862–78.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122–47.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1990). Mechanisms of moral disengagement. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, theologies, states of mind* (pp. 161–91). Cambridge: Cambridge University Press.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45–103). Hillsdale, NJ: Lawrence Erlbaum.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1–46). New York: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462–79.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126*, 925–45.

Baumeister, R. F. (1987). How the self became a problem: A psychological review of historical research. *Journal of Personality and Social Psychology, 43*, 425–36.

Baumeister, R. F. (1994). The crystallization of discontent in the process of major life change. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 281–97). Washington, DC: American Psychological Association.

Baumeister, R. F. (2005a). *The cultural animal: Human nature, meaning, and social life.* New York: Oxford University Press.

Baumeister, R. F. (2005b). Self and volition. In W. R. Miller & H. D. Delaney (Eds.), *Judeo-Christian perspectives on psychology: Human nature, motivation, and change.* Washington, DC: American Psychological Association.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation.* New York: Academic Press.

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development, 37*(887–907).

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience, 8*, 1458–63.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*, 568–86.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In R. M. Lerner (Ed.), *Handbook of child psychology* (5th ed., Vol. 1: Theoretical models of human development, pp. 993–1028). New York: Wiley.

Broughton, J. M. (1986). The psychology, history, and ideology of the self. In K. Larsen (Ed.), *Dialectics and ideology in psychology* (pp. 128–63). Norwood, NJ: Ablex.

Corno, L. (1993). The best-laid plans: Modern conceptions of volition and educational research. *Educational Researcher, 22*(2), 14–22.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum Press.

Diaz, R. M., Neal, C. J., & Vachio, A. (1991). Maternal teaching in the zone of proximal development: A comparison of low- and high-risk dyads. *Merrill Palmer Quarterly, 37*, 83–108.

Diaz, R. M., Winsler, A., Atencio, D. J., & Harbors, K. (1992). Mediation of self-regulation through the use of private speech. *International Journal of Education & Mediated Learning,* 155–67.

DiClemente, C. C. (2003). *Addiction and change: How addictions develop and addicted people recover.* New York: Guilford Press.

DiClemente, C. C., Marinilli, A. S., Singh, M., & Bellino, L. E. (2001). The role of feedback in the process of health behavior change. *American Journal of Health Behavior, 25*, 217–27.

Drakeford, J. W. (1967). *Integrity therapy.* Nashville, TN: Broadman.

Fingarette, H. (1988). *Heavy drinking: The myth of alcoholism as a disease.* Berkeley, CA: University of California Press.

Ford, D. (1987). *Humans as self-constructing living systems: A developmental perspective on behavior and personality.* Hillsdale, NJ: Lawrence Erlbaum.

Frankl, V. E. (1963). *Man's search for meaning.* Boston: Beacon Press.

Frankl, V. E. (1969). *The will to meaning.* New York: World Publishing.

Gollwitzer, P. M. (1999). Implementation intentions: Simple effects of simple plans. *American Psychologist, 54*, 493–503.

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in action: The importance of implementation intentions. *Personality and Social Psychology Review, 2*, 124–36.

Gottlieb, G. (1992). *Individual development and evolution: The genesis of novel behavior.* New York: Oxford University Press.

Grolnick, W. S., & Ryan, R. M. (1989). Parent styles associated with children's self-regulation and competence in school. *Journal of Educational Psychology, 81*, 143–54.

Harré, R., & Secord, P. F. (1972). *The explanation of social behavior.* Oxford: Basil Blackwell.

Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual Review of Clinical Psychology, 1*, 91–111.

Howard, G. S. (1986). *Dare we develop a human science?* Notre Dame, IN: Academic Publications.

Howard, G. S., & Conway, C. G. (1986). Can there be an empirical science of volitional action? *American Psychologist, 41*, 1241–51.

James, W. (1899). *Talks to teachers on psychology and to students on some of life's ideals.* New York: Holt.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E.

Nisbett, S. Valins, & B. Weiner (Eds.), *Attributions: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

Kanfer, F. H. (1970). Self-regulation: Research, issues, and speculation. In C. Neuringer & J. L. Michael (Eds.), *Behavior modification in clinical psychology* (pp. 178–220). New York: Appleton-Century-Crofts.

Koob, G. F. (2005). The neurobiology of addiction: A hedonic Calvinist perspective. In W. R. Miller & K. M. Carroll (Eds.), *Rethinking substance abuse: What the science shows, and what we should do about it* (pp. 25–45). New York: Guilford Press.

Kopp, C. B. (1982). The antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*, 99–214.

Kopp, C. B. (1987). The growth of self-regulation: Parents and children. In N. Eisenberg (Ed.), *Contemporary topics in developmental psychology* (pp. 34–55). New York: Wiley.

Kopp, C. B. (1991). Young children's progression to self-regulation. In M. Bullock (Ed.), *Contributions to human development* (Vol. 22: The development of intentional action: Cognitive, motivational, and interactive processes, pp. 38–54). Basel: Karger.

Kuhl, J. (1985). Volitional mediators of cognitive-behavior consistency: Self-regulatory processes and action versus state orientation. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 101–28). Berlin: Springer-Verlag.

Kuhl, J., & Beckmann, J. (Eds.). (1985). *Action control: From cognition to behavior.* Berlin: Springer-Verlag.

Lander, N. R., & Nelson, D. (2005). *The integrity model of existential psychotherapy in working with the "difficult patient."* New York: Routledge.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Lazarick, D. L., Fishbein, S. S., Loriello, M. A., & Howard, G. S. (1988). Practical investigators of volition. *Journal of Counseling Psychology, 35*, 15–26.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Logan, F. A. (1993). Animal learning and motivation and addictive drugs. *Psychological Reports, 73*, 291–306.

Lutkenhaus, P., & Bullock, M. (1991). The development of volitional skills. In M. Bullock (Ed.), *Contributions to human development* (Vol. 22: The development of intentional action: Cognitive, motivational, and interactive processes, pp. 14–23). Basel: Karger.

Mahoney, M. J., & Thoresen, C. E. (1974). *Self-control: Power to the person.* Monterey, CA: Brooks/Cole.

Maisto, S. A., Clifford, P. R., Stout, R. L., & Davis, C. M. (2006). Drinking in the year after treatment as a predictor of three-year drinking outcomes. Journal of Studies on Alcohol, 67, 823–32.

Malle, B. F., & Nelson, S. E. (2003). Judging *mens rea:* The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law, 21*, 563–80.

Marlatt, G. A., & Donovan, D. M. (Eds.). (2005). *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors* (2nd ed.). New York: Guilford Press.

McRae, R. R., & Costa, P. T., Jr. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd ed.). New York: Guilford Press.

Mead, G. H. (1934). *Mind, self, and society from the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

Meichenbaum, D. (1977). *Cognitive-behavior modification*. New York: Plenum Press.

Meichenbaum, D., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. *Journal of Personality and Social Psychology, 34*, 943–50.

Miller, G. F. (1997). Protean primates: The evolution of adaptive unpredictability in competition and courtship. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 312–40). Cambridge, UK: Cambridge University Press.

Miller, W. R., Benefield, R. G., & Tonigan, J. S. (1993). Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. *Journal of Consulting and Clinical Psychology, 61*, 455–61.

Miller, W. R., & Brown, J. M. (1991). Self-regulation as a conceptual basis for the prevention and treatment of addictive behaviours. In N. Heather, W. R. Miller, & J. Greeley (Eds.), *Self-control and the addictive behaviours* (pp. 3–79). Sydney: Maxwell Macmillan Publishing Australia.

Miller, W. R., & Carroll, K. M. (Eds.). (2006). *Rethinking substance abuse: What science shows and what we should do about it*. New York: Guilford Press.

Miller, W. R., & Delaney, H. D. (Eds.). (2005). *Judeo-Christian perspectives on psychology: Human nature, motivation and change.* Washington, DC: American Psychological Association.

Miller, W. R., & Muñoz, R. F. (2005). *Controlling your drinking.* New York: Guilford Press.

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd ed.). New York: Guilford Press.

Mowrer, O. H. (1983). *Leaves from many seasons: Selected papers.* New York: Praeger.

Mowrer, O. H., Vattano, A. J., et al. (1974). *Integrity groups: The loss and recovery of community.* Urbana, Illinois: Integrity Groups. The cover of the book literally says "and Others," without identifying them.

Mowrer, O. H., & Vattano, A. J. (1976). Integrity groups: A context for growth in honesty, responsibility, and involvement. *Journal of Applied Behavioral Sciences, 12*, 419–31.

Nuttin, J. R. (1987). The respective roles of cognition and motivation in behavioral dynamics, intention, and volition. In F. Halisch & J. Kuhl (Eds.), *Motivation, intention and volition* (pp. 156–171). Heidelberg: Springer-Verlag.

Perugini, M., & Bagozzi, R. P. (2004). The distinction between desires and intentions. *European Journal of Social Psychology, 34*, 69–84.

Pinker, S. (2002). *The blank slate: The modern denial of human nature.* New York: Penguin Books.

Premack, D. (1972). Mechanisms of self-control. In W. A. Hunt (Ed.), *Learning mechanisms in smoking* (pp. 107–23). Chicago: Aldine.

Prochaska, J. O. (1994). Strong and weak principles for progressing from precontemplation to action on the basis of twelve problem behaviors. *Health Psychology, 13*, 47–51.

Prochaska, J. O., & DiClemente, C. C. (1984). *The transtheoretical approach: Crossing traditional boundaries of therapy.* Homewood, Illinois: Dow/Jones Irwin.

Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol, 58*, 7–29.

Project MATCH Research Group. (1998). Matching alcoholism treatments to client heterogeneity: Project MATCH three-year drinking outcomes. *Alcoholism: Clinical and Experimental Research, 22*, 1300–11.

Riegel, K. (1976). The dialectics of human development. *American Psychologist, 31*, 689–700.

Rise, J., Thompson, M., & Verplanken, B. (2003). Measuring implementation intentions in the context of the theory of planned action. *Scandinavian Journal of Psychology, 44*, 87–95.

Rogoff, B., Turkanis, C., & Bartlett, L. (2001). *Learning together: Children and adults in a school community.* New York: Oxford University Press.

Rokeach, M. (1973). *The nature of human values.* New York: Free Press.

Ross, L., & Nisbett, R. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In N. Eisenberg (Ed.), *Handbook of child psychology* (5th ed., Vol. 3: Social, emotional and personality development, pp. 105–76). New York: Wiley.

Searle, J. (2001). *Rationality in action.* Cambridge, MA: MIT Press.

Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243–310). New York: Macmillan.

Spears, R., Jetten, J., & Scheepers, D. (2002). Distinctiveness and definition of collective self: A tripartite model. In A. Tesser, D. A. Stapel, & J. V. Wood (Eds.), *Self and motivation: Emerging psychological perspectives* (pp. 147–72). Washington, DC: American Psychological Association.

Taylor, C. (1991). *The ethics of authenticity.* Cambridge, MA: Harvard University Press.

Thoresen, C. E., & Mahoney, M. J. (1974). *Behavioral self-control.* New York: Holt, Rinehart & Winston.

Trevarthen, C. (1993). The self born in intersubjectivity: The psychology of an infant communicating. In U. Neisser (Ed.), *The perceived self: Ecological and interpersonal sources of self-knowledge* (pp. 121–73). New York: Cambridge University Press.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* New York: Cambridge University Press.

Wertsch, J. V., Tulviste, P., & Hagstrom, F. (1993). A sociocultural approach to agency. In E. R. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 336–56). New York: Oxford University Press.

Zhu, J. (2004). Intention and volition. *Canadian Journal of Philosophy, 34,* 175–94.

15 Willing Creation: The Yin and Yang of the Creative Life

Dean Keith Simonton

Free will presupposes the power to select among options. It presumes that one has the capacity to make choices. It means not just the direct experience of volition but also the realization of an act of will. It is not just a subjective state but rather the objective manifestation of the consequences of that state. This necessity is represented in the prototypical act of creation described in the third verse of the first chapter of the first book of the Bible: "And God said, 'Let there be light'; and there was light." But this is the story of a Creator who is conveniently omniscient and omnipotent. Will instantaneously becomes action. What about human creators? Can they also produce creations with in the same forthright act of will? Or are they, unlike the biblical divine being, less capable of converting aspiration to achievement? Are their accomplishments decoupled from their ambitions? Are human creators more the passive pawns of fate than the willful agents of creativity?

In thinking about this question, it soon becomes clear that human creativity represents something of a paradox. On the one hand, few areas of human behavior require so much will power. On the other hand, in few areas is the will so powerless. Speaking metaphorically, creativity appears to depend on two diametrically opposed forces: yin (passive) and yang (active). Let me explain.

CREATIVITY AS YIN

The biblical story portrays the Creator's creativity as active. Yet mythology often views human creativity in a quite contrary manner. This contrast is most conspicuous in the Greek myth of the Muses. As the story goes, Zeus, the reigning god in the Greek pantheon, fathered nine daughters, each of whom presided over a different domain of human attainment. More specifically, Muses were responsible for heroic or epic poetry, lyric and love poetry, sacred poetry, tragedy, comedy, music, dance, astronomy, and history. Each Muse was thought to provide a guiding spirit or source of inspiration for the mortal creator. This usage underlies several commonplace expressions. For instance, we may say that a novelist experiencing "writer's block" is waiting for her Muse. And we conclude that an artist has lost his Muse when he has run out of creative ideas. According to this conception, the Muse, like God, is the active agent of creativity, and the human creator must passively wait for some act of divine grace.

Admittedly, myths are not truths. Actual human creativity could be far more willful and deliberate. Yet the introspective reports of highly creative individuals imply that creativity is indeed more yin than yang (Hadamard, 1945; Ghiselin, 1952; Poincaré, 1921). Rather than a deliberate application of some inevitable logic or method, creative ideas often appear as flashes of insight or unexpected inspirations. The classic example is the Eureka experience by which Archimedes solved a major problem at the moment he immersed himself in a bathtub. These illuminations arrive when least expected—most often when engaged in some mundane activity unrelated to the problem at hand (Boden, 1991). In fact, these unanticipated discoveries or inventions typically occur when the creators have given up on solving a particular problem and thus have entered an "incubation period" in which they work on other problems that seem more amenable to solution (Wallas, 1926). In other words, creativity will happen only after creators stop trying to will their creativity to happen. A common explanation for this phenomenon is that by giving up volitional control, the creator can rely more on unconscious thought processes that cannot be directed by the will.

Another illustration comes from the dramatic event known as serendipity (Cannon, 1940; Roberts, 1989; Shapiro, 1986). This occurs when scientists make a discovery that was not what they were originally searching for. Looking for the solution to one problem, they chance upon a solution to a totally different problem. This is to be distinguished from "pseudoserendipity" in which someone accidentally stumbles upon the solution to a problem they have been explicitly trying to solve (Díaz de Chumaceiro, 1995). In true serendipity the creative idea emerged by chance, without any act of will, whereas in pseudoserendipity the creator was trying unsuccessfully to make something happen but managed to pull it off only by the intervention of luck.

The phenomenon of serendipity suggests that creation often cannot be willed because it is contingent on the intrusion of chance. Yet it may also happen that creativity is independent of human volition because it totally determined by forces external to the individual. A case in point is the phenomenon of multiple discovery and invention (Lamb & Easton, 1984). This is the occasion when two or more individuals come up with the same contribution independently and often simultaneously. Multiples run in the hundreds and include such classic examples as the theory of evolution by natural selection (Darwin and Wallace) and the calculus (Newton and Leibniz). Sociologists and cultural anthropologists have long argued that these events prove the irrelevance of the individual creator (e.g., Kroeber, 1917; Merton, 1961b; Ogburn & Thomas, 1922). Discoveries and inventions are not the products of creators but rather the manifestations of sociocultural determinism—the zeitgeist. At a particular time and place in history, a given idea becomes absolutely inevitable. At other times and places, the same act of creativity would be impossible. Although there is reason to believe that sociocultural determinists have tremendously overstated their case (Simonton, 2004), the explanatory principle may apply to a subset of creative ideas. In those instances, at least, personal desires, ambitions, goals, or aspirations would have no causal significance.

Naturally, the preceding examples seem more anecdotal than scientific. Certainly they cannot have the same status as actual scientific investigations. Yet a considerable body of empirical research also lends support to the inference that the act of creation is not always a volitional act. For instance, laboratory experiments on insightful problem solving indicate that the success of the incubation period largely depends on the "opportunistic assimilation" of unanticipated environmental stimuli (Seifert, Meyer, Davidson, Patalano, & Yaniv, 1995). Hence, the solution to a problem may be contingent on a fortuitous external event that happens to prime a fruitful series of associations. Furthermore, psychometric inquiries have shown that creative individuals suffer from reduced latent inhibition and negative priming (Carson, Peterson, & Higgins, 2003; Eysenck, 1993). That is, the creative mind is less adept at filtering out irrelevant information, a deficiency that implies a certain lack of willful control over the contents of thought. Those who can exert the most control over the intellect are those who are less likely to be creative.

It is also worth noting that several theories of the creative process appear to underplay the importance of volition in the act of creation. This is certainly apparent in the classic psychoanalytic conception of "regression in the service of the ego." Whereas the ego is the seat of the will, creativity depends more on the relinquishing of control to descend into the realm of primary-process imagery and unconscious thought (Ochse, 1989; Suler, 1980). Indeed, this view has received indirect endorsement in research indicating that the solution of insight problems is less likely when individuals are expected to engage in conscious, deliberate thinking (Schooler & Melcher, 1995).

Another set of theories with comparable implications are those based on the Darwinian idea of blind variation and selective retention (Simonton, 1999a). Because ideational variations must be to some extent blind, the creator cannot will a creation into being but rather must engage in some variety of haphazard trial-and-error process. Excellent illustrations can be seen in Picasso's sketches for *Guernica*, one of his most famous paintings (Simonton, in press). Rather than a systematic progression toward the end state, such as we would expect from an act of willful honing of the initial compositional sketches, what is observed is a chaotic search of diverse possibilities, Picasso sometimes getting "hotter" and other times "colder" with respect to the final product.

One final set of research findings are germane to this issue: the behavior genetics of creativity. Creative individuals can be distinguished from noncreative individuals on a large number of cognitive and dispositional characteristics (Feist, 1998). Moreover, most of these traits have reasonably high hereditability coefficients (Simonton, 1999b). For instance, a certain threshold level of intelligence—often specified as IQ 120—is required for a person to exhibit truly exceptional creativity (Simonton, 1994). Yet intelligence is clearly subject to genetic inheritance. Another trait is psychoticism, as measured by the Eysenck Personality Questionnaire (Eysenck, 1995). Not only is this personal attribute correlated with creativity, but it also has a substantial genetic loading (Eysenck, 1995). Hence, it is conceivable that a substantial proportion of the human population will not have the genetic constitution requisite for the display of exceptional creativity. This possibility is especially likely if creativity requires the joint inheritance of a large number of genetic traits (i.e., multidimensional and multiplicative; see Simonton, 1999b). For this subset of the population, there will be no way even if there is a will. These are the people who want desperately to build a better mousetrap, find the cure for cancer, or write a great novel—and yet never manage to do so. Only those who receive sufficient genetic gifts enjoy the potential to convert volition to action.

CREATIVITY AS YANG

So far it would seem that creators are somewhat passive individuals. They must wait for inspiration to come from some external source, whether it be chance or the zeitgeist. The more control they try to assert over the process, the farther they will be from their goal. Creativity cannot be forced, even among those with the creative capacity. And a great many people may not even have the ability to display creativity to any substantial degree. Even so, there is another side to creativity, a side where yang is more prominent than yin.

For instance, although creativity is partly born, as indicated by behavior genetics, it is also made. In other words, creative individuals are the products of nurture, not just nature. Furthermore, some aspects of nurture presume the operation of volition. Most conspicuous is the fact that creativity requires the deliberate and laborious acquisition of domain-specific expertise. This requirement has been expressed as the "10-year rule" (Ericsson, 1996). Specifically, a person cannot reach world-class levels of creative performance without first devoting a full decade to time-intensive study and practice. Although there is evidence that highly gifted people can take less than 10 years to master the necessary knowledge and skills (Simonton, 1991, 2000), it remains the case that this preparatory phase of the career cannot be skipped altogether. Moreover, this preparatory period demands considerable focused concentration, self-discipline, and determination. In other words, preparation presupposes an act of willpower extended over a long span of time. Those who lack volitional strength and consistency will never advance beyond the status of an amateur or dilettante.

Another illustration comes from later in the career, when the creative potential established through expertise acquisition becomes actualized as overt creative behavior. A career of creativity is seldom without obstacles and frustrations. On the contrary, creators may find that it is not always easy to get work published or exhibited or performed, and even when a work passes that hurdle, the creator may have to face scathing reviews from critics, unfavorable responses from the public, or negative reactions from colleagues. It is rare for any creator to experience a continuous series of successes unpunctuated by any failures. This means that a premium is placed on those individuals who have the motivational fortitude to stick it out no matter how severe the setbacks. As a consequence, drive, persistence, determination, and tenacity have long been identified as critical attributes of highly creative individuals (e.g., Cox, 1926; Roe, 1953).

In addition, some of the phenomena showing the role of yin, when examined more closely, also betray the place of yang. To start with, consider the incubation period that precedes the moment of creative illumination. Although this period requires that the will be disengaged from the creative process, the will was very active prior to the incubation period. Prior to incubation was a preparation period in which the creator is actively trying to solve a particular problem (Wallas, 1926). Not only is the individual deliberately choosing a specific problem to solve, but even after giving up on the problem, and entering the incubation phase, he or she is willfully holding the problem on the "back burner." The conscious mind may be directed at other tasks, but the unconscious mind remains committed to the unanswered question. In fact, sometimes the incubation period will last years, even decades, before the creator's preoccupation is rewarded by a major insight. That can happen only because the creator intends to solve the problem sooner or later. For instance, it took Einstein about a decade to solve a contradiction in theoretical physics that he first noticed when he was 16 years old.

The multiples phenomenon offers another case in point. There exist substantial individual differences in the frequency that scientists and inventors are involved in such duplicate discoveries and inventions (Simonton, 2004). These individual differences can be tied to acts of will. In the first place, highly productive individuals are more likely to get involved in multiples than those much less prolific (Merton, 1961a; Simonton, 1979). That is, by chance alone, those who generate the most ideas are more prone to produce ideas that others have generated. Insofar as prolific output is the outcome of willful determination, then so are multiples. Hence, a scientist or inventor could lower the odds of participating in multiples simply by lowering productivity. Naturally, this suggestion is not very useful because few creators would curtail their contributions to a scientific domain simply because they do not want to share the credit with other creators. Yet a second consideration is more interesting. Persons are most likely to get involved in multiples to the extent that they are active in "hot" research areas with many competitors trying to make the same breakthrough (Simonton, 2004). As a consequence, an individual can lower the likelihood of multiple involvement merely by shifting to less fashionable areas of research, a decision many scientists actually make in practice (Hagstrom, 1974). This decision also lessens the chance that one might get preempted by some rival and thus have done all one's work in vain.

THE YIN AND YANG OF CREATIVITY

Taken altogether, creativity's relation to volition appears to be characterized by both contingency and independence. In some respects, creativity seems determined by outside forces beyond the individual's control—forces like chance and the zeitgeist—whereas in other respects, the very act of creation seems heavily contingent on an act of will. Now I want to close this essay with a final observation that somewhat merges these two contrary viewpoints. That observation concerns the nature of creativity when it attains the highest levels. When we speak of creative genius—of high-caliber creators like Newton, Dostoyevsky, Michelangelo, and Beethoven—it becomes difficult to separate the yin and the yang. The very lives of these individuals are propelled forward by a sense of destiny, a feeling that they have some special mission, a notion that they have a goal in life that supersedes all others. This means that their drive is driven, that the self-actualization that underlies their creativity is a deterministic repercussion of who they are in the deepest portions of their being. Their willful output is not so much an act of will as a deliberate manifestation of their very identity, an identity that they can do nothing to change. Like *the* Creator, great human creators create because creativity is what defines who they are. So what are we to make of this obsession with mission, this compulsion toward creation? Is it passive or active? Yin or yang? I don't know. I'll let you decide.

REFERENCES

Boden, M. A. (1991). *The creative mind: Myths & mechanisms.* New York: Basic Books.

Cannon, W. B. (1940). The role of chance in discovery. *Scientific Monthly, 50,* 204–9.

Carson, S., Peterson, J. B., & Higgins, D. M. (2003). Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *Journal of Personality and Social Psychology, 85*, 499–506.

Cox, C. (1926). *The early mental traits of three hundred geniuses.* Stanford, CA: Stanford University Press.

Díaz de Chumaceiro, C. L. (1995). Serendipity or pseudoserendipity? Unexpected versus desired results. *Journal of Creative Behavior, 29*, 143–47.

Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry, 4*, 147–78.

Eysenck, H. J. (1995). *Genius: The natural history of creativity.* Cambridge, England: Cambridge University Press.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to expert performance: Empirical evidence from the arts and sciences, sports, and games* (pp. 1–50). Mahwah, NJ: Erlbaum.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review, 2*, 290–309.

Ghiselin, B. (Ed.). (1952). *The creative process: A symposium*. Berkeley: University of California Press.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton, NJ: Princeton University Press.

Hagstrom, W. O. (1974). Competition in science. *American Sociological Review, 39*, 1–18.

Kroeber, A. L. (1917). The superorganic. *American Anthropologist, 19*, 163–214.

Lamb, D., & Easton, S. M. (1984). *Multiple discovery.* Avebury, England: Avebury.

Merton, R. K. (1961a). The role of genius in scientific advance. *New Scientist, 12*, 306–8.

Merton, R. K. (1961b). Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society, 105*, 470–86.

Ochse, R. (1989). A new look at primary process thinking and its relation to inspiration. *New Ideas in Psychology, 7*, 315–30.

Ogburn, W. K., & Thomas, D. (1922). Are inventions inevitable? A note on social evolution. *Political Science Quarterly, 37*, 83–93.

Poincaré, H. (1921). *The foundations of science: Science and hypothesis, the value of science, science and method* (G. B. Halstead, Trans.). New York: Science Press.

Roberts, R. M. (1989). *Serendipity: Accidental discoveries in science.* New York: Wiley.

Roe, A. (1953). *The making of a scientist*. New York: Dodd, Mead.

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97–133). Cambridge, MA: MIT Press.

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 65–124). Cambridge, MA: MIT Press.

Shapiro, G. (1986). *A skeleton in the darkroom: Stories of serendipity in science.* San Francisco: Harper & Row.

Simonton, D. K. (1979). Multiple discovery and invention: Zeitgeist, genius, or chance? *Journal of Personality and Social Psychology, 37,* 1603–16.

Simonton, D. K. (1991). Emergence and realization of genius: The lives and works of 120 classical composers. *Journal of Personality and Social Psychology, 61,* 829–40.

Simonton, D. K. (1994). *Greatness: Who makes history and why.* New York: Guilford Press.

Simonton, D. K. (1999a). *Origins of genius: Darwinian perspectives on creativity.* New York: Oxford University Press.

Simonton, D. K. (1999b). Talent and its development: An emergenic and epigenetic model. *Psychological Review, 106,* 435–57.

Simonton, D. K. (2000). Creative development as acquired expertise: Theoretical issues and an empirical test. *Developmental Review, 20,* 283–318.

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist.* Cambridge, England: Cambridge University Press.

Simonton, D. K. (in press). The creative imagination in Picasso's *Guernica* sketches: Monotonic improvements or nonmonotonic variants? *Creativity Research Journal.*

Suler, J. R. (1980). Primary process thinking and creativity. *Psychological Bulletin, 88,* 144–65.

Wallas, G. (1926). *The art of thought.* New York: Harcourt, Brace.

16 Free Will Requires Determinism

John Baer

The will is infinite, and the execution confined…
the desire is boundless, and the act a slave to limit.
—William Shakespeare, *Troilus and Cressida*, III. ii.75–77

In the textbook of my first introductory psychology class there was, in chapter 1, a simple equation:

Behavior = Heredity + Environment + ?

The question mark, it was suggested, might include such things as chance, acts of God, and free will. Chance would make regular reappearances in my psychological studies. Acts of God and free will, on the other hand, were rarely if ever mentioned. One could assume only that if these were real influences on behavior, they were beyond the ken of psychology, and psychology would therefore go about its business as if they didn't matter, assuming (and demonstrating, as psychology's explanatory and predictive powers grew) that they had at most a very limited influence on human behavior.

Until then, I had never really thought much about free will, and this equation troubled me enough to set in motion a now almost 4-decade-old struggle to make sense of it. The redoutable B. F. Skinner (the most influential living psychologist when I was an undergraduate) rather directly, and the rest of psychology somewhat less directly, seemed to deny the possibility of free will. But even those who might have been more amenable to a belief in free will, such as the humanistic psychologists of the late 1960s, seemed to suggest (in agreement with the equation) that free will belonged, together with chance and possible acts of God, in the unpredictable, nondeterministic part of any account of human behavior.

Since the quantum revolution, few scientists believe that the world is totally deterministic. But if, at any moment, anything could happen—which is another way of saying that the past has no control over the present and future—then any kind of prediction or control, or even understanding, would be impossible. Psychology can proceed only to the extent that the universe is deterministic. So psychologists rather naturally attend to those aspects of human behavior that follow (or that they assume follow) discernible cause-and-effect logic.

To avoid possible confusion, I should make clear the definition of determinism I will be using in this chapter. Determinism is a theory or belief that events, including acts of the will, occurrences in nature, and social or psychological phenomena, are causally determined by preceding events and natural laws. Determinism assumes that all events in the universe, including all the things that happen in human minds, follow laws of causality.

It is hard to see how free will could be part of a deterministic universe (or could exist in that part of the universe—the deterministic part—that psychologists try to understand). Philosophers use the term *incompatibilism* to describe the belief that free will and determinism are incompatible. Incompatibilism claims that if the universe is deterministic, then we can't have free will. Unfortunately, there's no easy way out of this situation, because you can't get back free will just by arguing against determinism. The opposite of determinism is indeterminism, and (as I will discuss shortly) indeterminism is *totally* incompatible with any notion that we are in control of what we do.

I alluded to quantum mechanics earlier and to the chance, probabilistic, and ultimately indeterminate nature of the subatomic world of quarks and of the strange forces that affect these almost infinitely tiny building blocks of the universe. Determinism can adjust for quantum uncertainty because at the level we can observe phenomena—the level where we live, the world of things we can perceive without splitting atoms—all of these tiny chance effects tend to disappear, sort of like the way different parts of an algebra equation often cancel each other out. But it's nonetheless true that, over time, those tiny subatomic indeterminacies can add up and result in truly random events in the macroscopic world in which we live. And this accumulation of random events, though its overall effect is small, makes it impossible, in principle, for anyone to predict the future with complete accuracy, no matter how all-knowing that person might be.

But adding chance to the mix doesn't rescue free will from determinism. One can argue that because our brains are made of subatomic particles that are subject to chance events, and because this means that it is impossible to determine precisely what we will think or how we will behave, determinism has been defeated. This is correct: At least to a small degree, according to our current best scientific understanding, our universe is indeed indeterminate.

That said, it's important to understand that the distinction between determinism and indeterminism is really not so much an either–or question as it is a matter of degree. At the level we are able to notice effects—at the level of readily observable events—the little bit of randomness that quantum mechanics tells us about makes almost no discernible difference in our lives, and its effects on our will and on the decisions we make are at most slight. There may indeed be rare moments of seeming randomness that result, down the road, in huge differences in our lives, as was exemplified in the wonderful movie *Sliding Doors* a few years ago. In that movie, the main character (played by Gwyneth Paltrow) rushes to catch a train as the doors are closing, and we are shown the very different ways her life plays out both in a world in which she makes the train and one in which she just misses it, a kind of chance event that has widespread ramifications.

But even if we could overturn determinism by saying we live in a universe determined not by natural laws but by chance, that doesn't help us at all on the question of free will. If my behavior is the result of chance, I'm not in control; chance is (or because chance isn't really an agent, one might simply say that nothing is in control—but, either way, it's clear that I am not in control to the extent that events are "determined" by chance). If the thought that your behavior is the complex result of some combination of all the genetic and environmental influences that have touched you is disturbing and causes you to question free will, then it's unlikely that the idea that your behavior is the result of purely chance events that are totally out of your control will provide any reassurance. We don't want our lives to be a kind of cosmic game of craps, with control given over to the roll of the dice. To whatever extent randomness actually rules in the universe, it effectively excludes that much possibility of control by any of us. So a totally random, indeterminate universe is most definitely a universe without the possibility of free will, and we can be thankful that we do not live in such a universe. If we are to have free will, it will have to come from a deterministic universe, one in which there are causes of behavior.

HOW WE MAKE DECISIONS

In the one-person play *Defending the Caveman*, the following line, which I am paraphrasing, tends to evoke an initial stunned silence, and then some moans and hisses: "When arguing, women aren't limited by the rules of rational thinking."

The actor then asks people if they agree with this provocative remark. After all the women and most of the men in the audience murmur their disagreement, he offers to turn it around: "When arguing, women *are* limited by the rules of rational thinking." Does that make it better? Not really—in fact, I hope it helps us remember that rational thinking is only one way that we think and reason and make sound judgments and decisions. So not being constrained by the rules of rational thinking is a positive thing, not an insult.

Humans are at least *somewhat* rational creatures, however. We are not Spock-like; our rationality has limits, and it is both impeded and, often, assisted by emotions and other human attributes that are not strictly rational. Our reasoning powers are complex and many. The important idea here is a rather obvious one—that we are able to reason, both rationally and in other ways, and our reasoning is part of who we are and how we make decisions. This seems so obvious that one might wonder why I've even bothered to mention it, but it's a necessary piece of my answer to the free will question.

On the TV show *Who Wants to be a Millionaire?*—or, for that matter, on a multiple-choice exam that I might give in one of my classes—as long as one gets the right answer, it doesn't really matter whether one actually knew the answer or just made a lucky guess because the outcome is the same. But we can also understand that there *is* nonetheless a very important difference. In the same way, choosing a wise course of action because one had good reasons for choosing it is different from choosing a lucky course of action, even though the wise choice and the lucky choice might be the same and have the same effects. We feel that choosing because we have reasons for choosing makes a decision more *our* choice than choosing by flipping a coin.

Humans reason—both rationally and in other ways—and our reasoning influences the decisions we make. When I was in college, psychedelic drugs were very popular. People who used them found their reasoning followed different tracks and resulted in different decisions. Whether their reasoning was impaired or heightened needn't concern us, nor do we need to judge whether the decisions people made on drugs were better or worse; the important, if obvious, point is that these decisions were *different* from what they would have been had drugs not been involved. Our reasoning and decision-making abilities, whether operating optimally or suboptimally, greatly influence our thoughts and behaviors.

FREE WILL UNDER DETERMINISM

It doesn't solve the free will problem, but it's a step along the way to acknowledge that our reasoning and decision making do influence what we do. It's part of who we are, it's part of how we make decisions, and it is a major factor in determining whether we do one thing or another.

But in a deterministic universe, our reasoning, although it influences our decisions and actions, is also itself determined by things that have come before—by what we know, by the people we've known, by some genetic factors that influence how we go about reasoning, and by many other such things that have gone into making us who we are. So we do use reasoning, and the reasoning we do *does* help determine what we decide to do from moment to moment; but that reasoning ability and the ways we exercise it remain, in a deterministic universe, subject to natural, physical laws, even if those influences are so well hidden in a zillion brain synapses that we could never see exactly how it works or predict in advance what decision is sure to result in any given situation.

Does the fact that who I am at any moment in time commit me to acting in a certain way eliminate my free will? I suppose that depends on what we mean by free will. When Martin Luther declared his disagreement with certain Church teachings, he said, "Here I stand. I can do no other." Did he mean that he had no control over his actions? Of course not. If he had had reason to believe that by holding back, or by making a different set of objections, he might somehow end all human suffering, then I'm fairly confident that he would have done something different. But, given the situation that he found himself in, and given the kind of person he was with the kinds of beliefs and understandings that he had, he—the person he was—could do no other. Doing anything else would have been untrue to himself.

Free will means having the power to do different things, and to choose to do what makes most sense at the moment. It means we will choose what it is most in our natures at any moment to do. Are those choices caused? Certainly. They are caused by a combination of our natures—who we are at that moment, something that has been shaped by both genes and experiences—and the actual constraints of the situation in which we find ourselves. We can know that whatever a person is doing at any moment, it is in accord with that person's nature and with the situation in which he finds himself, however he came to have his particular nature and to be in that particular situation.

Each of us has many courses of action that are possible in the sense that they are within our power—we *could* do them if we choose to do so—but we act only in ways that accord with our natures, at any moment in time and in any given situation, by making the particular choices we make. Free will doesn't mean doing things that make no sense. Free will means that your thinking, reasoning, emotions, personality, memories, goals, decision-making strategies, and everything else that makes you who you are actually *matter.* Are our lives and choices therefore predictable? Well, given even small amounts of quantum uncertainty, no, not in perfect detail; but, in a larger sense, yes. All of us are, in general, fairly predictable, which is a good thing if you think about the amount of predictable cooperation that is necessary for us to do things like drive cars on roads used by other drivers. And most of the decisions we make seem to make sense, and are in that sense predictable in terms of who we are and what our goals and desires and skills and attitudes and beliefs happen to be. But can

I know what those decisions will be without going through the kinds of reasoning, emoting, thinking, and other behaviors that constitute the way I make decisions? No, it's simply impossible. No one will ever be able to have that kind of foreknowledge.

So we do have free will in a deterministic universe. Indeterminism, on the other hand, makes free will impossible, because random events by definition cannot be under our control. To the extent that determinism is true, we humans do indeed have something that we all innately feel and believe that we have: free will. In this most important sense, determinism makes free will possible and meaningful.

Some might argue that this isn't *truly* free will. It is true that no one has created himself ex nihilo, and if we trace back the cause-and-effect chain to its beginning, one can argue that it began even before one's birth. If one wants the kind of free will that denies cause and effect, a free will that would disengage one's past from the present, then one is seeking either randomness or supernatural intervention, not free will. But if who a person is (her personality, cognitive abilities, beliefs, ideas, emotions, memories, wishes, thinking styles, etc.) is to have power over what she does—and isn't this what we really mean by free will?—then the only kind of free will that is coherent is deterministic free will.

Determinism makes free will possible. It also makes psychology possible. If psychological events were not determined—caused—by antecedent events, psychology could make no sense. We have a lot for which to thank determinism, both as psychologists and as free will–possessing humans.

ACKNOWLEDGMENTS

I have discussed free will for many years, with almost anyone who would share her ideas or listen to mine, and there isn't enough space here to thank all those many friends and acquaintances individually. I'm sure many of those people will recognize their influence in what I've written here, and I also wish to thank those whose attempts to persuade me to different viewpoints have left less obvious traces.

A few people have been especially influential in shaping my thinking. Dan Dennett, whose own chapter also appears in this book, had a very profound influence on my thinking when I first encountered his arguments about free will 2 decades ago (and I strongly recommend his excellent book *Elbow Room: The Varieties of Free Will Worth Wanting,* as well as his more recent books). I am sure he will recognize many of his ideas in this short chapter.

More recently, during a sabbatical semester spent at Yale, I had the good fortune to audit a course on early modern philosophy taught by Keith DeRose. He was very tolerant of my many questions about free will and directed me

to literature that often challenged my views. I also participated in a 6-week seminar at Yale entitled "God, Free Will, and the Problem of Evil" under the direction of Scott Ragland. Although (like Keith) Scott would be unlikely to agree with much of what I've written here, his ideas (and those of the other participants in the seminar) helped shape my ideas in many ways, and I am deeply appreciative of those influences.

17 The Fear of Determinism

Steven Pinker

This essay is about the common fear that an understanding of human nature in biological terms is a form of *determinism* in the sense that is opposed to *free will* in introductory philosophy courses. The fear of determinism is captured in a following limerick:

There was a young man who said: "Damn!
It grieves me to think that I am
Predestined to move
In a circumscribed groove:
In fact, not a bus, but a tram.

In the traditional conception of a ghost in the machine, our bodies are inhabited by a self or a soul that chooses the behavior to be executed by the body. These choices are not compelled by some prior physical event, like one billiard ball smacking into another and sending it into a corner pocket. The idea that our behavior is caused by the physiological activity of a genetically shaped brain would seem to refute the traditional view. It would make our behavior an automatic consequence of molecules in motion and leave no room for an uncaused behavior-chooser.

One fear of determinism is a gaping existential anxiety: that deep down we are not in control of our own choices. All our brooding and agonizing over the right thing to do is pointless, it would seem, because everything has already been preordained by the state of our brains. If you suffer from this anxiety, I suggest the following experiment. For the next few days, don't bother deliberating over your actions. It's a waste of time, after all; they have already been determined. Shoot from the hip, live for the moment, and if it feels good do it. No, I am not seriously suggesting that you try this! But a moment's reflection on what would happen if you *did* try to give up making decisions should serve as a Valium for the existential anxiety. The experience of choosing is not a fiction, regardless of how the brain works. It is a real neural process, with the obvious function of selecting behavior according to its foreseeable consequences. It responds to information from the senses, including the exhortations of other people. You cannot step outside it or let it go on without you because it *is* you. If the most ironclad form of determinism is real, you could not do anything about it anyway, because your anxiety about determinism, and how you would deal with it, would also be determined. It is the existential fear of determinism that is the real waste of time.

A more practical fear of determinism is captured in a saying by A. A. Milne: "No doubt Jack the Ripper excused himself on the grounds that it was human nature." The fear is that an understanding of human nature seems to eat away at the notion of personal responsibility. In the traditional view, the self or soul, having chosen what to do, takes responsibility when things turn out badly. As with the desk of Harry Truman, the buck stops here. But when we attribute an action to a person's brain, genes, or evolutionary history, it seems that we no longer hold the individual accountable. Biology becomes the perfect alibi, the get-out-of-jail-free card, the ultimate doctor's excuse note. As we have seen, this accusation has been made by the religious and cultural right, who want to preserve the soul, and the academic left, who want to preserve a "we" who can construct our own futures though in circumstances not of our own choosing.

Why is the notion of free will so closely tied to the notion of responsibility, and why is biology thought to threaten both? Here is the logic. We blame people for an evil act or bad decision only when they intended the consequences and could have chosen otherwise. We don't convict a hunter who shoots a friend he has mistaken for a deer, or the chauffeur who drove John F. Kennedy into the line of fire, because they could not foresee and did not intend the outcome of their actions. We show mercy to the victim of torture who betrays a comrade, to a delirious patient who lashes out at a nurse, or to a madman who strikes someone he believes to be a ferocious animal, because we feel they are not in command of their faculties. We don't put a small child on trial if he causes a death, nor do we try an animal or an inanimate object, because we believe them to be constitutionally incapable of making an informed choice.

A biology of human nature would seem to admit more and more people into the ranks of the blameless. A murderer may not literally be a raving lunatic, but our newfangled tools might pick up a shrunken amygdala or a hypometabolism in his frontal lobes or a defective gene for MAO oxidase, which renders him just as out of control. Or perhaps a test from the cognitive psychology lab will show that he has chronically limited foresight, rendering him oblivious to consequences, or that he has a defective theory of mind, making him incapable of appreciating the suffering of others. After all, if there is no ghost in the machine, *something* in the criminal's hardware must set him apart from the majority of people, those who would not hurt or kill in the same circumstances. Pretty soon we will find this something, and, it is feared, murderers will be excused from criminal punishment as surely as we now excuse madmen and small children.

Even worse, biology may show that we are *all* blameless. Evolutionary theory says that the ultimate rationale for our motives is that they perpetuated our ancestors' genes in the environment in which we evolved. Since none of us are aware of that rationale, none of us can be blamed for pursuing it, any more than we blame the mental patient who thinks he is subduing a mad dog but really is attacking a nurse. We scratch our heads when we learn of ancient customs that punished the soulless: the Hebrew rule of stoning an ox to death if it killed a man, the Athenian practice of putting an ax on trial if it injured a man (and hurling it over the city wall if found guilty), a medieval French case in which a sow was sentenced to be mangled for having mauled a child, and the whipping and burial of a church bell in 1685 for having assisted French heretics.[1] But evolutionary biologists insist we are not fundamentally different from animals, and molecular geneticists and neuroscientists insist we are not fundamentally different from inanimate matter. If people are soulless, why is it not just as silly to punish people? Shouldn't we heed the creationists, who say that if you teach children they are animals they will behave like animals? Should we go even further than the National Rifle Association bumper sticker—GUNS DON'T KILL; PEOPLE KILL—and say that not even people kill, because people are just as mechanical as guns?

These concerns are by no means academic. Cognitive neuroscientists are sometimes approached by criminal defense lawyers hoping that a wayward pixel on a brain scan might exonerate their client (a scenario that is wittily played out in Richard Dooling's novel *Brain Storm*). When a team of geneticists found a rare gene that predisposed the men in one family to violent outbursts, a lawyer for an unrelated murder defendant argued that his client might have those genes, too. If so, the lawyer argued, "his actions may not have been a product of total free will."[2] When Randy Thornhill and Craig Palmer argued that rape is a consequence of male reproductive strategies, another lawyer contemplated using their theory to defend rape suspects.[3] (Insert your favorite lawyer joke here.) Biologically sophisticated legal scholars, such as Owen Jones, have argued that a "rape gene" defense would almost certainly fail, but the general threat remains that biological explanations will be used to exonerate wrongdoers.[4] Is this the bright future promised by the sciences of human nature—it wasn't me, it was my amygdala? Darwin made me do it? The genes ate my homework?

People hoping that an uncaused soul might rescue personal responsibility are in for a disappointment. In *Elbow Room: The Varieties of Free Will Worth Wanting*, the philosopher Dan Dennett points out that the last thing we want in a soul is freedom to do anything it desires.[5] If behavior were chosen by an utterly free will, then we *really* couldn't hold people responsible for their actions. That entity would not be deterred by the threat of punishment, or be ashamed by the prospect of opprobrium, or even feel the twinge of guilt that might inhibit a sinful temptation in the future, because it could always choose to defy those causes of behavior. We could not hope to reduce evil acts by enacting moral and legal codes, because a free agent, floating in a different plane from the arrows of cause and effect, would be unaffected by the codes. Morality and law would be pointless. We could punish a wrongdoer, but it would be sheer spite, because it could have no predictable effect on the future behavior of the wrongdoer or of other people aware of the punishment.

On the other hand, if the soul *is* predictably affected by the prospect of esteem and shame or reward and punishment, it is no longer truly free, because it is compelled (at least probabilistically) to respect those contingencies. Whatever converts standards of responsibility into changes in the likelihood of behavior—such as the rule "If the community would think you're a boorish cad for doing X, don't do X"—can be programmed into an algorithm and implemented in neural hardware. The soul is superfluous.

Defensive scientists sometimes try to deflect the charge of determinism by pointing out that behavior is never perfectly predictable but always probabilistic, even in the dreams of the hardest-headed materialists. (In the heyday of Skinner's behaviorism, his students formulated the Harvard Law of Animal Behavior: "Under controlled experimental conditions of temperature, time, lighting, feeding, and training, the organism will behave as it damned well pleases.") Even identical twins reared together, who share all of their genes and most of their environment, are not identical in personality and behavior, just highly similar. Perhaps the brain amplifies random events at the molecular or quantum level. Perhaps brains are nonlinear dynamical systems subject to unpredictable chaos. Or perhaps the intertwined influences of genes and environment are so complicated that no mortal will ever trace them out with enough precision to predict behavior exactly.

The less-than-perfect predictability of behavior certainly gives the lie to the cliché that the sciences of human nature are "deterministic" in the mathematical sense. But it doesn't succeed in allaying the fear that science is eroding the concept of free will and personal responsibility. It is cold comfort to be told that a man's genes (or his brain or his evolutionary history) made him 99% likely to kill his landlady as opposed to 100%. Sure, the behavior was not strictly preordained, but why should the 1% chance of his having done otherwise suddenly make the guy "responsible"? In fact, there is *no* probability value that, by itself, ushers responsibility back in. One can always think that there is a 50% chance some molecules in Raskolnikov's brain went thisaway, compelling him to commit the murder, and a 50% chance they went thataway, compelling him not to. We still have nothing like free will, and no concept of responsibility that promises to reduce harmful acts. Philosophers call it Hume's Fork: "Either our actions are determined, in which case we are not responsible for them, or they are the result of random events, in which case we are not responsible for them."

People who hope that a ban on biological explanations might restore personal responsibility are in for the biggest disappointment of all. The most risible pretexts for bad behavior in recent decades have come not from biological determinism but from *environmental* determinism: the abuse excuse, the Twinkie defense, black rage, pornography poisoning, societal sickness, media violence, rock lyrics, and different cultural mores (recently used by one lawyer to defend a Gypsy con artist and by another to defend a Canadian Indian woman who murdered her boyfriend).[6] Just in the week I wrote this paragraph, two new examples appeared in the newspapers. One is from a clinical psychologist who "seeks out a dialogue" with repeat murderers to help them win mitigation, clemency, or an appeal. It manages to pack the doctrine of the blank slate, the doctrine of the noble savage, the moralistic fallacy, and environmental determinism into a single passage:

> Most people don't commit horrendous crimes without profoundly damaging things happening to them. It isn't that monsters are being born right and left. It's that children are being born right and left and are being subjected to horrible things. As a consequence, they end up doing horrible things. And I would much rather live in that world than in a world where monsters are just born.[7]

The other is about a social work student in Manhattan:

> Tiffany F. Goldberg, a 25-year-old from Madison, Wis., was struck on the head with a chunk of concrete by a stranger this month. Afterward, she expressed concern for her attacker, speculating that he must have had a troubled childhood.
>
> Graduate students in social work at Columbia called Ms. Goldberg's attitude consistent with their outlook on violence. "Society is into blaming individuals," said Kristen Miller, 27, one of the students. "Violence is intergenerationally transmitted."[8]

Evolutionary psychologists are commonly chided for "excusing" men's promiscuity with the theory that a wandering eye in our ancestors was rewarded with a greater number of descendants. They can take heart from a recent biography that said Bruce Springsteen's "self-doubts made him frequently seek out the sympathy of groupies,"[9] a book review that said Woody Allen's sexual indiscretions "originated in trauma" and an "abusive" relationship with his mother,[10] and

Hillary Clinton's explanation of her husband's libido in her infamous interview in *Talk:*

> He was so young, barely 4, when he was scarred by abuse that he can't even take it out and look at it. There was terrible conflict between his mother and grandmother. A psychologist once told me that for a boy being in the middle of a conflict between two women is the worst possible situation. There is always the desire to please each one.[11]

Mrs. Clinton was raked by the pundits for trying to excuse her husband's sexual escapades, though she said not a word about brains, genes, or evolution. The logic of the condemnation seems to be, if someone tries to explain an act as an effect of some cause, the explainer is saying that the act was not freely chosen and that the actor cannot be held responsible.

Environmental determinism is so common that a genre of satire has grown around it. In a *New Yorker* cartoon, a woman on a witness stand says, "True, my husband beat me because of his childhood; but I murdered him because of mine." In the comic strip *Non Sequitur,* the directory of a mental health clinic reads, "1st Floor: Mother's Fault. 2nd Floor: Father's Fault. 3rd Floor: Society's Fault." And who can forget the Jets in *West Side Story,* who explained to the local police sergeant, "We're depraved on accounta we're deprived"?

> Dear kindly Sergeant Krupke,
> You gotta understand,
> It's just our bringing up-ke,
> That gets us out of hand.
> Our mothers all are junkies,
> Our fathers all are drunks.
> Golly Moses, naturally we're punks!

Something has gone terribly wrong. It is a confusion of *explanation* with *exculpation.* Contrary to what is implied by critics of biological *and* environmental theories of the causes of behavior, to explain behavior is not to exonerate the behaver. Hillary Clinton may have advanced the dumbest explanation in the history of psychobabble, but she does not deserve the charge of trying to excuse the president's behavior. (A *New York Times* story described Mr. Clinton's response to people's criticism of his wife: "'I have not made any excuses for what was inexcusable, and neither has she, believe me,' he said, arching his eyebrows for emphasis.")[12]

If behavior is not utterly random, it will have some explanation; if behavior *were* utterly random, we couldn't hold the person responsible in any case. So if we *ever* hold people responsible for their behavior, it will have to be in spite of any causal explanation we feel is warranted, whether it invokes genes, brains, evolution, media images, self-doubt, bringing up-ke, or being raised by bickering women. The difference between *explaining* behavior and *excusing* it is captured in the saying "To understand is not to forgive," and has been stressed in different ways by many philosophers, including Hume, Kant, and Sartre.[13] Most philosophers believe that unless a person was literally coerced (that is, someone held a gun to his head), we should consider his actions to have been freely chosen, even if they were caused by events inside his skull.

But *how* can we have both explanation, with its requirement of lawful causation, and responsibility, with its requirement of free choice? To have them both we don't need to resolve the ancient and perhaps unresolvable antinomy between free will and determinism. We only have to think clearly about what we want the notion of responsibility to achieve. Whatever may be its inherent abstract worth, responsibility has an eminently practical function: deterring harmful behavior. When we say that we hold someone responsible for a wrongful act, we expect him to punish himself—by compensating the victim, acquiescing to humiliation, incurring penalties, or expressing credible remorse—and we reserve the right to punish him ourselves. Unless a person is willing to suffer some unpleasant (and hence deterring) consequence, claims of responsibility are hollow. Richard Nixon was ridiculed when he bowed to pressure and finally "took responsibility" for the Watergate burglary but did not accept any costs such as apologizing, resigning, or firing his aides.

One reason to hold someone responsible is to deter the person from committing similar acts in the future. But that cannot be the whole story, because it is different only in degree from the contingencies of punishment used by behaviorists to modify the behavior of animals. In a social, language-using, reasoning organism, the policy can also deter similar acts by *other* organisms who learn of the contingencies and control their behavior so as not to incur the penalties. That is the ultimate reason we feel compelled to punish elderly Nazi war criminals, even though there is little danger that they would perpetrate another holocaust if we let them die in their beds in Bolivia. By holding them responsible—that is, by publicly enforcing a policy of rooting out and punishing evil wherever and whenever it occurs—we hope to deter others from committing comparable evils in the future.

This is not to say that the concept of responsibility is a recommendation by policy wonks for preventing the largest number of harmful acts at the least cost. Even if experts had determined that punishing a Nazi would prevent no future atrocities, or that we could save more lives by diverting the manpower to catching drunk drivers, we would still want to bring Nazis to justice. The demand for responsibility can come from a burning sense of just deserts, not only from literal calculations of how best to deter particular acts.

But punishment even in the pure sense of just deserts is *ultimately* a policy for deterrence. It follows from a paradox inherent to the logic of deterrence: Though the *threat* of punishment can deter behavior, if the behavior does take place the punishment serves no purpose other than pure sadism or an illogical desire to make the threat credible retroactively. "It won't bring the victim back," say the opponents of capital punishment, but that can be said about *any* form of punishment. If we start the movie at the point at which a punishment is to be carried out, it looks like spite, because it is costly to the punisher and inflicts harm on the punishee without doing anyone any immediate good. In the middle decades of the 20th century, the paradox of punishment and the rise of psychology and psychiatry led some intellectuals to argue that criminal punishment is a holdover from barbaric times and should be replaced by therapy and rehabilitation. The position was clear in the titles of books like George Bernard Shaw's *The Crime of Imprisonment* and the psychiatrist Karl Menninger's *The Crime of Punishment*. It was also articulated by leading jurists such as William O. Douglas, William Brennan, Earl Warren, and David Bazelon. These radical Krupkeists did not suffer from a fear of determinism; they welcomed it with open arms.

Few people today argue that criminal punishment is obsolete, even if they recognize that (other than incapacitating some habitual criminals) it is pointless in the short run. That is because if we ever did calculate the short-term effects in deciding whether to punish, potential wrongdoers could *anticipate* that calculation and factor it into their behavior. They could predict that we would not find it worthwhile to punish them once it was too late to prevent the crime, and could act with impunity, calling our bluff. The only solution is to adopt a resolute policy of punishing wrongdoers *regardless* of the immediate effects. If one is genuinely not bluffing about the threat of punishment, there is no bluff to call. As Oliver Wendell Holmes explained, "If I were having a philosophical talk with a man I was going to have hanged (or electrocuted) I should say, 'I don't doubt that your act was inevitable for you but to make it more avoidable by others we propose to sacrifice you to the common good. You may regard yourself as a soldier dying for your country if you like. But the law must keep its promises.'"[14] This promise-keeping underlies the policy of applying justice "as a matter of principle," regardless of the immediate costs or even of consistency with common sense. If a death-row inmate attempts suicide, we speed him to the emergency ward, struggle to resuscitate him, give him the best modern medicine to help him recuperate, and kill him. We do it as part of a policy that closes off all possibilities to "cheat justice."

Capital punishment is a vivid illustration of the paradoxical logic of deterrence, but the logic applies to lesser criminal punishments, to personal acts of revenge, and to intangible social penalties like ostracism and scorn. Evolutionary psychologists and game theorists have argued that the deterrence paradox led to the evolution of the emotions that undergird a desire for justice: the implacable need for retribution, the burning feeling that an evil act knocks the universe out of balance and can be canceled only by a commensurate punishment. People who are emotionally driven to retaliate against those who cross them, even at a cost to themselves, are more credible adversaries and less likely to be exploited.[15] Many judicial theorists argue that criminal law is simply a controlled implementation of the human desire for retribution, designed to keep it from escalating into cycles of vendetta. The Victorian jurist James Stephen said that "the criminal law bears the same relation to the urge for revenge as marriage does to the sexual urge."[16]

Religious conceptions of sin and responsibility simply extend this lever by implying that any wrongdoing that is undiscovered or unpunished by one's fellows will be discovered and punished by God. Martin Daly and Margo Wilson sum up the ultimate rationale of our intuitions about responsibility and godly retribution:

> From the perspective of evolutionary psychology, this almost mystical and seemingly irreducible sort of moral imperative is the output of a mental mechanism with a straightforward adaptive function: to reckon justice and administer punishment by a calculus which ensures that violators reap no advantage from their misdeeds. The enormous volume of mystico-religious bafflegab about atonement and penance and divine justice and the like is the attribution to higher, detached authority of what is actually a mundane, pragmatic matter: discouraging self-interested competitive acts by reducing their profitability to nil.[17]

The deterrence paradox also underlies the part of the logic of responsibility that makes us expand or contract it when we learn about a person's mental state. Modern societies do not just pick whatever policy is most effective at deterring wrongdoers. For example, if one's only value was to reduce crime, one could always make the punishments for it especially cruel, as most societies did until recently. One could convict people on the basis of an accusation, a guilty manner, or a forced confession. One could execute the entire family of a criminal, or his entire clan or village. One could say to one's adversaries, as Vito Corleone said to the heads of the other crime families in *The Godfather,* "I'm a superstitious man. And if some unlucky accident should befall my son, if my son is struck by a bolt of lightning, I will blame some of the people here."

The reason these practices strike us as barbaric is that they inflict more harm than is necessary to deter evil in the future. As the political writer Harold Laski said, "Civilization means, above all, an unwillingness to inflict unnecessary pain." The problem with broad-spectrum deterrents is that they catch innocent people in their nets, people who could not have been deterred from committing an undesirable act to start with (such as the kin of the man who pulled the trigger, or a bystander during a lightning storm that kills the Godfather's son). Since punishment of these innocents could not possibly deter other people like them, the harm has no compensating benefit even in the long run, and we consider it unjustified. We seek to fine-tune our policy of punishment so that it applies only to people who *could have been* deterred by it. They are the ones we "hold responsible," the ones we feel "deserve" the punishment.

A fine-tuned deterrence policy explains why we exempt certain harm-causers from punishment. We don't punish those who were unaware that their acts would lead to harm, because such a policy would do nothing to prevent similar acts by them or by others in the future. (Chauffeurs cannot be deterred from driving a president into the line of fire if they have no way of knowing there will be a line of fire.) We don't apply criminal punishment to the delirious, the insane, small children, animals, or inanimate objects, because we judge that they—and entities similar to them—lack the cognitive apparatus that could be informed of the policy and could inhibit behavior accordingly. We exempt these entities from responsibility not because they follow predictable laws of biology while everyone else follows mysterious not-laws of free will. We exempt them because, unlike most adults, they lack a functioning brain system that can respond to public contingencies of punishment.

And this explains why the usual exemptions from responsibility should *not* be granted to all males or all abuse victims or all of humanity, even when we think we can explain what led them to act as they did. The explanations may help us understand the parts of the brain that made a behavior tempting, but they say nothing about the *other* parts of the brain (primarily in the prefrontal cortex) that could have inhibited the behavior by anticipating how the community would respond to it. We are that community, and our major lever of influence consists in appealing to that inhibitory brain system. Why should we discard our lever on the system for inhibition just because we are coming to understand the system for temptation? If you believe we shouldn't, that is enough to hold people responsible for their actions—without appealing to a will, a soul, a self, or any other ghost in the machine.

This argument parallels a long-running debate about the most blatant example of a psychological explanation that nullifies responsibility, the insanity defense.[18] Many legal systems in the English-speaking world follow the 19th-century M'Naughten rule:

> . . . the jurors ought to be told in all cases that every man is to be presumed to be sane, and to possess a sufficient degree of reason to be responsible for his crimes, until the contrary be proved to their satisfaction; and that, to establish a defense on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was laboring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing, or, if he did know it, that he did not know he was doing what was wrong.

This is an excellent characterization of a person who cannot be deterred. If someone is too addled to know that an act would harm someone, he cannot be inhibited by the injunction "Don't harm people, or else!" The M'Naughten rule aims to forgo spiteful punishment—retribution that harms the perpetrator with no hope of deterring him or people similar to him.

The insanity defense achieved its present notoriety, with dueling rent-a-shrinks and ingenious abuse excuses, when it was expanded from a practical test of whether the cognitive system responding to deterrence is working to the more nebulous tests of what can be said to have produced the behavior. In the 1954 *Durham* decision, Bazelon invoked "the science of psychiatry" and "the science of psychology" to create a new basis for the insanity defense:

> The rule we now hold is simply that an accused is not criminally responsible if his unlawful act was the product of mental disease or mental defect.

Unless one believes that ordinary acts are chosen by a ghost in the machine, *all* acts are products of cognitive and emotional systems in the brain. Criminal acts are relatively rare—if everyone in a defendant's shoes acted as he did, the law against what he did would be repealed—so heinous acts will often be products of a brain system that is in *some* way different from the norm, and the behavior can be construed as "a product of mental disease or mental defect." The *Durham* decision and similar insanity rules, by distinguishing behavior that is a product of a brain condition from behavior that is something else, threatens to turn every advance in our understanding of the mind into an erosion of responsibility.

Now, some discoveries about the mind and brain really could have an impact on our attitudes toward responsibility—but they may call for expanding the domain of responsibility, not contracting it. Suppose desires that sometimes culminate in the harassment and battering of women are present in many men. Does that really mean that men should be punished more leniently for such crimes, because they can't help it? Or does it mean they should be punished more surely and severely, because that is the best way to counteract a strong or widespread urge? Suppose a vicious psychopath is found to have a defective sense of sympathy, which makes it harder for him to appreciate the suffering of his victims. Should we mitigate the punishment because he has diminished capacity? Or should we make the punishment more sure and severe to teach him a lesson in the only language he understands?

Why do people's intuitions go in opposite directions—both "If he has trouble controlling himself, he should be punished more leniently" and "If he has trouble controlling himself, he should be punished more severely"? It goes back to the deterrence paradox. Suppose some people need a threat of 1 lash with a wet noodle to deter them from parking in front of a fire hydrant. Suppose people with a bad gene, a bad brain, or a bad childhood need the threat of 10 lashes. A policy that punishes illegal parkers with nine lashes will cause unnecessary suffering and not solve the problem: 9 lashes is more than necessary to deter ordinary people and less than necessary to deter defective people. Only a penalty of 10 lashes can reduce both illegal parking and lashing: Everyone will be deterred, no one will block hydrants, and no one will get whipped. So, paradoxically, the two extreme policies (harsh punishment and no punishment) are defensible and the intermediate ones are not. Of course, people's deterrence thresholds in real life aren't pinned at just two values but are broadly distributed (1 lash for some people, 2 for others, and so on), so many intermediate levels of punishment will be defensible, depending on how one weighs the benefits of deterring wrongdoing against the costs of inflicting harm.

Even for those who are *completely* undeterrable, because of frontal lobe damage, genes for psychopathy, or any other putative cause, we do not have to allow lawyers to loose them on the rest of us. We already have a mechanism for those likely to harm themselves or others but who do not respond to the carrots and sticks of the criminal justice system: involuntary civil commitment, in which we trade off some guarantees of civil liberties against the security of being protected from likely predators. In all these decisions, the sciences of human nature can help estimate the distribution of deterrabilities, but they cannot weight the conflicting values of avoiding the greatest amount of unnecessary punishment and preventing the greatest amount of future wrongdoing.[19]

I do not claim to have solved the problem of free will, only to have shown that we don't need to solve it to preserve personal responsibility in the face of an increasing understanding of the causes of behavior. Nor do I argue that deterrence is the only way to encourage virtue, just that we should recognize it as the active ingredient that makes responsibility worth keeping. Most of all, I hope I have dispelled two fallacies that have allowed the sciences of human nature to sow unnecessary fear. The first fallacy is that biological explanations corrode responsibility in a way that environmental explanations do not. The second fallacy is that causal explanations (both biological and environmental) corrode responsibility in a way that a belief in an uncaused will or soul does not.

NOTES

1. Kaplan, 1973, p. 10.
2. E. Felsenthal, "Man's genes made him kill, his lawyers claim," *Wall Street Journal*, November 15, 1994. The defense was unsuccessful: see "Mobley v. The State," Supreme Court of Georgia, March 17, 1995, 265 Ga. 292, 455 S.E.2d 61.
3. "Lawyers may use genetics study in rape defense," *National Post* (Canada), January 22, 2000, p. A8.
4. Jones, 2000; Jones, 1999.
5. Dennett, 1984. See also Kane, 1998; Nozick, 1981, pp. 317–362; Ridley, 2000; Staddon, 1999.

6. Dershowitz, 1994; J. Ellement, "Alleged con man's defense: 'Different' mores," *Boston Globe*, February 25, 1999; N. Hall, "Metis woman avoids jail term for killing her husband," *National Post* (Canada), January 20, 1999.

7. B. English, "David Lisak seeks out a dialogue with murderers," *Boston Globe*, July 27, 2000.

8. M. Williams, "Social work in the city: Rewards and risks," *New York Times*, July 30, 2000.

9. S. Morse, Review of C. Sandford's *Springsteen point blank*, *Boston Globe*, November 19, 1999.

10. M. Udovich, Review of M. Meade's *The unruly life of Woody Allen*, *New York Times*, March 5, 2000.

11. L. Franks, Interview with Hillary Clinton, *Talk*, August 1999.

12. K. Q. Seelye, "Clintons try to quell debate over interview," *New York Times*, August 5, 1999.

13. Dennett, 1984; Kane, 1998; Nozick, 1981, pp. 317–62; Ridley, 2000; Staddon, 1999.

14. Quoted in Kaplan, 1973, p. 16.

15. Daly & Wilson, 1988; Frank, 1988; Pinker, 1997; Schelling, 1960.

16. Quoted in Kaplan, 1973, p. 29.

17. Daly & Wilson, 1988, p. 256.

18. Dershowitz, 1994; Faigman, 1999; Kaplan, 1973; Kirwin, 1997.

19. Rice, 1997.

REFERENCES

Daly, M., & Wilson, M. (1988). *Homicide*. Hawthorne, N.Y.: Aldine de Gruyter.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting.* Cambridge, Mass.: MIT Press.

Dershowitz, A. M. (1994). *The abuse excuse.* Boston: Little, Brown.

Faigman, D. L. (1999). *Legal alchemy: The use and misuse of science in the law*. New York: W. H. Freeman.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions.* New York: Norton.

Jones, O. (2000). Reconsidering rape. *National Law Journal*(February 21), A21.

Jones, O. D. (1999). Sex, culture, and the biology of rape: Toward explanation and prevention. *California Law Review, 87*, 827–942.

Kane, R. (1998). *The significance of free will*. New York: Oxford University Press.

Kaplan, J. (1973). *Criminal justice: Introductory cases and materials*. Mineola, NY: The Foundation Press.

Kirwin, B. R. (1997). *The mad, the bad, and the innocent: The criminal mind on trial.* Boston: Little, Brown.

Nozick, R. (1981). *Philosophical explanations.* Cambridge, Mass.: Harvard University Press.

Pinker, S. (1997). *How the mind works.* New York: Norton.

Rice, M. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist, 52*, 414–23.

Ridley, M. (2000). *Genome: The autobiography of a species in 23 chapters*. New York: HarperCollins.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, Mass.: Harvard University Press.

Staddon, J. R. (1999). On responsibility in science and law. In E. Paul, F. Miller & J. Paul (Eds.), *Responsibility* (Vol. 16). New York: Cambridge University Press.

18 Psychology and Free Will: A Commentary

Alfred R. Mele

In discussing the preceding chapters, I concentrate on themes that link many of them. In the first section, I provide some conceptual background on free will. The second section takes up some questions about determinism and laws of nature raised in the preceding chapters. In it, I suggest, among other things, that some contributors to this volume who seem to be disagreeing about whether psychology presupposes determinism may not actually be doing so. Instead, they may understand the word *determinism* differently. In the third section, I connect free will to consciousness, taking my lead from several of the preceding chapters. The fourth section takes up Benjamin Libet's work on free will and consciousness. That work is discussed in six of the preceding chapters, and Azim Shariff and coauthors (this volume) report that "almost all of the works involved in the deluge of anti–free will arguments have referenced" Libet: His work merits special attention here. In a recent volume entitled *Does Consciousness Cause Behavior?*, after briefly reviewing debates on their topic in the 19th and early 20th centuries, the editors write, "the wide promulgation of two new lines of genuinely scientific . . . evidence has seized the philosophical and scientific imagination and again brought the whole question to the forefront of intellectual debate" (Pockett et al. 2006, p. 1). They then identify Libet and Daniel Wegner as the sources of these two new lines of evidence. Wegner's work, including *The Illusion of Conscious Will* (2002), has received a lot of attention (in the present volume and elsewhere). In the fifth section, I consider the light that Wegner's contribution to this volume sheds on the "the illusion of conscious will" and on free will. The final section wraps things up.

FREE WILL: SOME PHILOSOPHICAL BACKGROUND

Free will may be defined as the power or ability to act freely (Mele 2006, pp. 16–17). But what is it to act freely? Familiar philosophical answers fall into two groups: compatibilist and incompatibilist. Compatibilism and incompatibilism are theses about the conceptual relationship between free action and determinism. *Determinism*, according to the standard philosophical definition of the term, is the thesis that a complete statement of the laws of nature together with a complete description of the condition of the entire universe at any point in time logically entails a complete description of the condition of the entire universe at any other point in time. (You and I are parts of the universe, and a description of what we are doing at this moment is part of a complete description of the universe at this moment.) *Compatibilism* is the thesis that free action is compatible with the truth of determinism. Because they attend to what contemporary physics tells us, the overwhelming majority of contemporary compatibilists do not believe that determinism (as defined here) is true; but they do believe that even if it were true, that would not preclude our being able to act freely. *Incompatibilism* is the thesis that free action is incompatible with the truth of determinism. In the incompatibilist group, most answers to the question what it is to act freely come from libertarians. *Libertarianism* is the conjunction of incompatibilism and the thesis that some people sometimes act freely. Some incompatibilists argue that no one acts freely. They argue that even the falsity of determinism creates no place for free action. (The chapters in this volume by Dweck and Molden and by Nichols provide useful references to philosophical work on the positions I have sketched.)

The compatibilist thesis often sounds strange to nonspecialists. When people first encounter the pair of expressions *free will* and *determinism*, they tend to get the impression that the two ideas are defined in opposition to each other, that they are mutually exclusive by definition. This is one reason that it is useful to think of free will as the power to act freely and to regard acting freely as the more basic notion—that is, as a notion in terms of which free will is to be defined. Consider the following conversation between two police officers who have a notoriously stingy friend named Stan.

ANN: Stan gave $20 to a homeless man today.
BILL: Why? Did he hold a gun to Stan's head?
ANN: No, Stan freely gave him the money.

Surely, Ann and Bill do not need to have an opinion about whether determinism (as defined above) is true to have this conversation. If what Ann says is true—that is, if Stan freely gave away $20—and free will is the power to act freely, then Stan has free will (or had it at that time). Even if *free will* is typically opposed to *determinism* in ordinary speech, *he freely did it* seems not to be. And even if *he freely did it* were typically opposed to determinism in ordinary speech, that would settle nothing. After all, in ordinary speech, deductive reasoning seems to be defined as reasoning from the general to the particular, and that certainly would only jokingly be said to constitute an objection to a logician's definition of deduction (according to which "Ann is a police officer; Bill is a police officer; therefore Ann and Bill are police officers" is a valid deductive argument).

Compatibilist theories of free action emphasize a distinction between deterministic causation and compulsion. If determinism is true, then my eating a banana for breakfast today and my working on this chapter today were deterministically caused; and so were a certain compulsive hand-washer's washing his hands dozens of times today, a certain delusional person's spending the day trying to contact God with his microwave oven, a certain addict's using his favorite drug while in the grip of an irresistible urge to do so, and a certain person's handing over money to gunmen who convincingly threatened to kill him if he refused. But there is an apparent difference. I am sane and free from addiction, and I received no death threats today. The basic compatibilist idea is (roughly) that when mentally healthy people act intentionally and rationally in the absence of compulsion and coercion, they act freely, and an action's being deterministically caused does not suffice for its being compelled or coerced.[1] (Chapters in this volume by Baer, Bandura, and Myers express a commitment to compatibilism.)

Many compatibilists have been concerned to accommodate the idea that, for example, if I freely spent the day working, I could have done something else instead. They grant that if determinism is true, then there is a sense in which people could never have done otherwise than they did: They could not have done otherwise in the sense that their doing otherwise is incompatible with the combination of the past and the laws of nature. But, these compatibilists say, the fact that a person never could have done otherwise in that sense is irrelevant to free action. What is relevant is that people who act freely are exercising a rational capacity of such a kind that if their situation had been different in any one of a variety of important ways, they would have responded to the difference with a different suitable action (Smith 2003). For example, although I spent the day working, I would have spent the day relaxing if someone had bet me $500 that I would not relax all day. This truth is consistent with determinism.[2] (Notice that if someone had made this bet with me, the past would have been different from what it actually was.) And it reinforces the distinction between deterministic causation and compulsion. Offer a compulsive hand-washer $500 not to wash his hands all day and see what happens.

Like compatibilists, libertarians tend to maintain that when mentally healthy people act intentionally in the absence of compulsion and coercion, they act freely, but libertarians insist that the deterministic causation of an action is incompatible with the action's being freely performed. (Recall that libertarians believe that determinism is false.) Some libertarian theories of free action assert that people never act freely unless some of their actions are indeterministically caused by immediate antecedents (Kane 1996). Whereas the laws that apply to deterministic causation are exceptionless, those that apply most directly to indeterministic causation are instead probabilistic.[3] Typically, events like *deciding* to help a stranded motorist—as distinct from the physical actions involved in actually helping—are counted as mental actions.[4] (As Miller and Atencio [this volume] observe, "deciding to do something is not the same as doing it.") Suppose that Ann's decision to help a stranded motorist is indeterministically caused by, among other things, her thinking that she should help. Because the causation is indeterministic, she might not have decided to help given exactly the same internal and external conditions. Some libertarians appeal to indeterministic causation to secure the possibility of doing otherwise that they require for free action.

Although some libertarians claim that free actions are uncaused, others—I will call them *causal libertarians*—hold that all actions are caused and that the *proximate* causes of basically free actions indeterministically cause them.[5] They maintain that free actions have proximate causes and that, in basic cases of free action, if a person freely does A at a time t, then in some possible, hypothetical scenario in which both the entire past up to t and the laws of nature are the same, he does not do A at t. Now, the proximate causes of actions, including actions that are decisions, are internal to agents. Even a driver's sudden decision to hit his brakes in an emergency situation is not proximately caused by events in the external world. Perception of whatever the source of the emergency happens to be—for example, a dog darting into traffic—is causally involved. And how the driver decides to react to what he sees depends on, among other things, his driving skills and habits, whether or not he is aware of what is happening directly behind him, and his preferences. A driver who likes driving over dogs and is always looking for opportunities to do that would probably react very differently than a normal person would. In light of the general point about the proximate causation of actions, typical causal brands of libertarianism encompass a commitment to what may be termed *agent-internal indeterminism*.[6]

What libertarians want that determinism precludes is not merely that agents have open to them more than one future that is compatible with the combination of the past and the laws of nature, but also that, on some occasions, which possible future becomes actual is in some sense and to some degree up to the agents. Causal libertarians want something that requires that people themselves be indeterministic in some suitable way—that some relevant things that happen under the skin are indeterministically caused by other such things. The focus is on mental events (or their neural correlates), as opposed, for example, to indeterministically caused muscle spasms—and, more specifically, on mental events that have a significant bearing on action (or the neural correlates of these events).

Quantum mechanics, according to leading interpretations, is indeterministic. But indeterminism at that level does not ensure that any human brains themselves sometimes operate indeterministically, much less that they sometimes operate indeterministically in ways appropriate for free action. One possibility, as David Hodgson reports, is that "in systems as hot, wet, and massive as neurons of the brain, quantum mechanical indeterminacies quickly cancel out, so that for all practical purposes determinism rules in the brain" (2002, p. 86). Another is that any indeterminism in the human brain is simply irrelevant to the production of actions. Empirical discoveries that either of these possibilities is an actuality would show that we do not have free will on some familiar libertarian conceptions of it.[7]

Many distinct libertarian and compatibilist theories about free will are in circulation. All of them have been challenged. Reviewing the major details of the debates is well beyond the scope of this chapter. My aim in this section has been to provide a theoretical context for the remainder of this chapter. In the same spirit, I conclude this section with brief comments on a pair of passages from Daniel Dennett's and John Bargh's contributions to this volume.

Dennett writes,

> If you are one of those who think that free will is only really free will if it springs from an immaterial soul that hovers happily in your brain, shooting arrows of decision into your motor cortex, then, given what you mean by free will, my view is that there is no free will at all. If, on the other hand, you think free will might be morally important without being supernatural, then my view is that free will is indeed real, but just not quite what you probably thought it was. (This volume, quoting from Dennett 2003, p. 222)

If one sets the bar for free will (that is, for the power or ability to act freely) ridiculously high, the thesis that people sometimes act freely should strike one as ridiculous. Where the bar should be set has long been a topic of philosophical debate. Although Dennett and I do not see entirely eye to eye about free will (see Dennett 2003 on Mele 1995, and Mele 2005 on Dennett 2003), I certainly agree with him that the only sensible place to look for it is in the natural order and that attention to moral responsibility helps to anchor free will in the real world. (On the point about anchoring, also see the chapters by Bandura and Pinker in this volume. For a less naturalistic perspective, see some of the remarks by Schooler in Shariff et al., this volume.)

A historical note with some conceptual implications is in order. John Bargh (this volume) writes, "It was St. Paul who discovered the notion of the individual will." This discovery, he contends, is present in Paul's report "that he knew what the good was, he wanted to do it, but could not always do it." Bargh adds, "In saying this, he introduced the key notion of individual control and responsibility for doing the right thing, and 'strength of will' as an important determinant of whether an individual will successfully do it instead of yielding to temptations of the flesh." In fact, the first 10 chapters of book 7 of Aristotle's *Nicomachean Ethics* are a detailed investigation of the phenomenon that Paul describes. And Aristotle is responding partly to Plato and Socrates. If the recognition that individuals sometimes struggle to resist temptation in the service of their judgments about what is good—sometimes successfully, sometimes not—is sufficient for having the "notion of individual control and responsibility," that notion was around long before Paul. And if, as Bargh seems to suggest, the notion just mentioned entails the notion of free will, the latter notion was around just as long. Whether the existence of individual control and the existence of individual moral responsibility jointly entail the existence of free will depends partly on what free will is. If Dennett and I are right, the entailment goes through.

DETERMINISM AND LAWS

Some of the contributors to this volume seem to disagree about the place of determinism in psychology. John Baer writes, "Determinism makes…psychology possible. If psychological events were not determined—caused—by antecedent events, psychology could make no sense." George Howard agrees: "If you want to be a scientist, you better be a determinist. Things are (and act) the way they are (and act) because something(s) caused them to be (or act) that way. It is a proper job for a scientist to find and document (via experimental studies) the cause-effect relations that form and guide human actions. Therefore, I am a determinist." However, Roy Baumeister reports that he resents "being told that as a scientist" he is "required to embrace total causal determinism," and he remarks that determinism "is contrary to our data, which almost invariably show probabilistic rather than deterministic causation." In a similar vein, Carol Dweck and Daniel Molden assert that "discovering predictability and lawfulness in human behavior does not imply determinism. We may measure certain personality factors and use our measures to predict people's behavior, but this does not mean that those factors . . . do not exert their influence in a probabilistic way."

Are Baer and Howard disagreeing with Baumeister, Dweck, and Molden, or are these two groups simply using the word *determinism* in two different ways? Suppose both groups were to agree to define *determinism* as I did in the first section, and suppose another technical term—*causalism*—were introduced and defined as the thesis that all psychological events, including intentional actions, are caused, either deterministically or indeterministically (probabilistically). Would the apparent disagreement between the two groups dissolve? There is reason to think that it would. Notice that Baer equates *determined* with *caused*. Howard seems to do the same in the passage I quoted. And, of course, causalism is compatible with the idea that "If psychological events were not…

caused…psychology could make no sense." By *determinism* Baer and Howard might not mean anything more demanding than causalism. And Baumeister and Dweck and Molden give no indication whatever that they would reject causalism, though they do seem to reject determinism, as I defined it.

Shaun Nichols (this volume) quotes the following from an article by John Bargh and Melissa Ferguson: "Psychologists studying higher mental processes should continue the scientific study of conscious processes but at the same time give appropriate attention to the deterministic philosophy that must underlie such analysis" (2000, p. 940). Nichols adds, "psychological determinism has been and will continue to be a vital assumption guiding research. And I'm inclined to think it's true.…[M]y allegiance…came from an abiding conviction that people's decisions have to have an explanation." What I called causalism is easily paired with the idea that all decisions have causal explanations, given that not all adequate causal explanations require that deterministic causes be at work. Whether decisions and other interesting events are *deterministically* caused is an empirical issue. Causalism is a sufficiently strong assumption for psychologists to proceed on.[8]

Dweck and Molden raise a question about how laws of nature are understood in a standard philosophical definition of *determinism*. Do "laws of human nature" count? What Albert Bandura (this volume) calls *epistemological reductionism* is relevant here. It "contends that the laws governing higher level psychosocial phenomena are ultimately reducible to the laws operating at the atomic and molecular levels." *If* the laws of physics are permanently in place shortly after the Big Bang, then *if* the universe is deterministic and devoid of nonphysical entities, it would seem that all that is needed for entailments of all future events and regularities is in place long before there are any living beings at all. Given these assumptions, the laws (regularities) of human nature would seem to be entailed by a complete description of the laws of physics and of the condition of the universe long before the advent of human beings. This upshot resembles the reductionism that Bandura has in mind. If the universe is not deterministic, the combination of the laws at the level of physics and the state of the universe at a given early time may leave it open to a significant extent what the psychological laws (regularities) will be.

I should add that to say that a universe is *not deterministic* (or *indeterministic*), as I use these terms, is just to say that determinism is not true of that universe. It certainly is not to say that causalism is false of it.

CONNECTING FREE WILL TO CONSCIOUSNESS

Whether compatibilism or incompatibilism is true is a conceptual question—not an empirical one.[9] (Notice that compatibilism, as I defined it, does not include the assertion that, in fact, people sometimes act freely.) But the following questions are empirical: (a) If incompatibilism is true, do people ever act freely? (b) If compatibilism is true, do people ever act freely? Scientists who weigh in on whether compatibilism is true or false do so as theorists, not as scientists. But scientists certainly can weigh in *as scientists* on questions (a) and (b).

How might a scientist who wishes to investigate free will without taking a stand on the compatibility of free will with determinism proceed? One way is to study phenomena that are associated with free will and can occur both in deterministic and in indeterministic universes: for example, self-regulation and decision making. Presumably, if physicists were to discover that determinism is true, we would not conclude that no one has ever successfully resisted temptation or that no one has ever made a decision (see chapters by Myers and Pinker, this volume). Incompatibilists would conclude that no one has ever *freely* done these things, but that is another matter. Some of the contributors to this volume proceed in the imagined way: creativity (Baumeister; Simonton) and self-control or self-regulation (Bandura; Baumeister; Howard; Miller & Atencio; Roediger et al.), for example, presuppose neither that determinism is false nor that it is true.

Dennett (this volume) mentions the idea that free action "depends on an agent's behavior being 'intelligible in terms of conscious purposes rather than in terms only of unconscious purposes.'" Presumably, like self-regulation and decision making, consciousness depends neither on the falsity nor on the truth of determinism. A lot of scientific work has been done on what place consciousness may or may not have in the production of actions. I discuss some of it later in this chapter.

Baumeister (this volume) writes, "if there are any genuine phenomena associated with the concept of free will, they most likely involve conscious choice. Such a view has to contend with the . . . belief that consciousness is a useless, feckless epiphenomenon, and that all behavior is guided by nonconscious processes." (On alleged epiphenomenalism about consciousness, see also, in this volume, the chapters by Bandura; Bargh; Kihlstrom; Myers; Roediger et al.; Shariff et al.; and Wegner). If all behavior were produced *only* by unconscious processes, and if conscious choices and their neural correlates were to play no role at all in producing any behavior at all, free will would be in dire straits. (Predictably, given what I have said already, I assume that conscious choices are wholly situated in the natural, causal order. More on this later.)

Some of our conscious choices or decisions are about things to do right away. Call them conscious *proximal* decisions. Others—conscious *distal* decisions—are about things to do later. A shy student who has been thinking about when it would be best to raise his hand to attract a lecturer's attention may decide to raise it *now.* And some time last week, after thinking about my various commitments and how I might efficiently honor them, I decided to start composing this chapter on Tuesday of this week. Sometimes, I make a note of my decisions on a "To Do" list, because I do not fully trust myself to remember them. But I was confident that I would remember my decision about this chapter, and I made no note. If I had not consciously remembered my plan to start working on my chapter this Tuesday, would I have done that? I strongly doubt it. And if I am right to doubt it, I am right to suggest that not all conscious experiences are epiphenomenal. After all, it certainly seems that my consciously remembering my plan helps to account for my starting to write my chapter on Tuesday. And my consciously making that plan—that is, my consciously deciding last week to start writing this chapter on Tuesday—certainly seems to have played a role in producing the conscious memory event.

To be sure, my starting to write my chapter on Tuesday without consciously remembering my plan is *possible.* For example, I might just coincidentally start it on Tuesday or a hypnotist might get me to start it then. But this possibility certainly does not entail that, in fact, my consciously remembering my plan did not help to account for the action at issue. (Consider an analogy. The last time I flew to Chicago, my boarding an Atlanta-to-Chicago flight was among the causes of my arriving in Chicago. My arriving in Chicago without boarding an Atlanta-to-Chicago flight definitely was possible: There were other ways for me to get to Chicago. But that certainly does not undermine the causal claim I made about my actual arrival there.)

Some scientific work on decision making is associated with skepticism about free will. The work I am most familiar with in this connection is on *proximal* decisions (and intentions), not distal ones. Some of that work is the topic of the next section.

FREE WILL AND CONSCIOUSNESS: LIBET'S WORK

I turn to Benjamin Libet's widely discussed work. A striking thesis of Libet, Gleason, et al. (1983) is that "the brain…'decides' to initiate or, at the least, prepare to initiate [certain actions] at a time before there is any reportable subjective awareness that such a decision has taken place" (p. 640; see also Libet 1985, p. 536).[10] Libet pointedly asserts, "If the 'act now' process is initiated unconsciously, then conscious free will is not doing it" (2001, p. 62; see also Libet, 2004, p. 136). As some contributors to this volume have observed, he also contends that there is about a 100 millisecond (ms) window of opportunity for free will to get involved and veto the brain's decision. Libet has many critics and many supporters. Some people follow him part of the way. They accept the thesis I quoted but reject the window of opportunity for free will as an illusion (Wegner 2002, p. 55).

In some of Libet's studies, subjects are regularly encouraged to flex their right wrists whenever they wish. In subjects who do not report any "preplanning" of their movements, electrical readings from the scalp—averaged over at least 40 flexings for each subject—show a shift in *readiness potentials* (RPs) beginning at about 550 ms before the time at which an electromyogram shows relevant muscular motion to begin (1985, pp. 529–30). These are "type II RPs" (p. 531). Subjects who are not regularly encouraged to aim for spontaneity or who report some preplanning produce RPs that begin about half a second earlier—"type I RPs." The same is true of subjects instructed to flex at a "preset" time (Libet et al. 1982, p. 325).

Subjects are also instructed to "recall...the spatial clock position of a revolving spot at the time of [their] initial awareness" (Libet 1985, p. 529) of something, *x*, that Libet variously describes as a "decision," "intention," "urge," "wanting," "will," or "wish" to move.[11] On the average, in the case of type II RPs, *RP onset* preceded what the subjects reported to be the time of their initial awareness of *x* (time *W*) by 350 ms. Time *W*, then, preceded the beginning of muscle motion by about 200 ms. I represent the results as follows:

LIBET'S RESULTS FOR TYPE II RPS

–550 ms	–200 ms	0 ms
RP onset	time *W*	muscle begins to move

(Libet finds independent evidence of a slight error in subjects' recall of the times at which they first become aware of sensations [1985, pp. 531, 534]. Correcting for it, time *W* is –150 ms.)

Henry Roediger and coauthors (this volume) write, "Clearly conscious intention cannot cause an action if a neural event that precedes and correlates with the action comes before conscious intention." This claim should strike readers as surprising. Consider the following claim: Clearly, the burning of a fuse cannot cause an explosion of a firecracker if a lighting of a fuse that precedes and correlates with the explosion comes before the burning of the fuse. Obviously, both the lighting of the fuse and the burning of the fuse are among the causes of the explosion. Other things being equal, if the fuse had not been lit—or if the lit fuse had stopped burning early—there would have been no explosion. Might it be that conscious proximal intentions to flex are part of the causal chain leading to the flexings of Libet's subjects? Also, is what precedes these conscious intentions anything more impressive than a potential cause of such intentions? More specifically, are these conscious intentions preceded by unconscious proximal decisions or intentions to flex? I set the stage for a discussion of these questions by raising another.

Is the brain activity registered by, say, the first 300 ms of type II RPs—call it *type 300 activity*—as tightly connected to subsequent flexings as lightings of firecracker fuses are to exploding firecrackers? In fact, we don't know. In the experiments that yield Libet's type II RPs, it is the muscle burst that triggers a computer to make a record of the preceding brain activity. In the absence of a muscle burst, there is no record of that activity. So, for all we know, there were many occasions on which type 300 activity occurred in Libet's subjects and there was no associated flexing.

Libet mentions that subjects encouraged to flex spontaneously "reported that during some of the trials a recallable conscious urge to act appeared but was 'aborted' or somehow suppressed before any actual movement occurred; in such cases the subject simply waited for another urge to appear, which, when consummated, constituted the actual event whose RP was recorded" (1985, p. 538). As he points out, "In the absence of the muscle's electrical signal when being activated, there was no trigger to initiate the computer's recording of any RP that may have preceded the veto" (2004, p. 141). So, for all we know, type 300 activity was present before the urges were suppressed. Because there was no muscle burst, no record was made of the brain activity.

Notice that it is *urges* that these subjects are said to report and "suppress." Might it be that type 300 activity is a potential cause of conscious urges to flex in Libet's subjects and that some subjects make no decision about when to flex—unconsciously or otherwise—until after the conscious urge emerges? And might it be that prior to the emergence of the conscious urge, subjects have no proximal intention to flex? Bandura (this volume) contends that it is important to distinguish urges from intentions and decisions, and I agree (Mele 1997, 2006, ch. 2). That our urges often are generated by processes of which we are not conscious is not at all surprising. And if we sometimes make effective decisions about whether or not to act on a conscious urge, so much the better for free will. Moreover, Libet's data do not show that subjects have unconscious proximal intentions to flex before they have conscious proximal intentions to flex. The data leave it open that what precedes these conscious intentions is a causal process that includes no unconscious proximal decisions or intentions to flex.

My point thus far is that Libet's data do not warrant either of the following claims: (L1) what happens earlier than, say,–200 ms in his subjects is causally sufficient for a muscle burst to occur at 0 ms; (L2) his subjects have proximal intentions to flex before they think they do. Some related issues merit attention.

Even if L2 is not warranted by Libet's data, his idea that we have unconscious proximal intentions should not be lightly dismissed. Some psychologists seem to view unconscious intentions as conceptually impossible (Wegner 2002, p. 18), and others disagree (Marcel 2003). I myself do not find the idea that there are unconscious proximal intentions at all disturbing (see Mele 2004). Such intentions may be at work when, for example, experienced drivers flip their turn indicators to signal for turns they are about to make. In a study in which subjects are instructed to flex whenever they feel like it *without also being instructed* to report after flexing on when they first became aware of an intention, urge, or, whatever, to flex, would they often be conscious of proximal intentions (or urges) to flex? Might unconscious proximal intentions to flex—and, more specifically, proximal intentions of which they are never conscious—be at work in producing flexings in the imagined scenario?

Imagine that the experiment I just sketched is conducted and it is discovered (somehow) that the subjects were never or rarely conscious of proximal urges or intentions to flex.[12] Could we legitimately infer that, in Libet's own experiment, conscious urges, decisions, and intentions had no effect on the flexings? No. One possibility is that some of Libet's subjects treat their initial consciousness of an urge to flex as a "go" signal (as suggested in Keller & Heckhausen 1990, p. 352). If they do, the "conscious urge" seemingly has a place in the causal process that issues in the flexing. Another possibility is that some subjects treat the conscious urge as what may be called a "decide" signal—a signal calling for them consciously to decide right then whether to flex right away or to wait a while. If that is so, and they consciously decide to flex and execute that decision, the conscious urge again seemingly has a place in the causal process, as does the conscious decision.

Perhaps it will be suggested that even if a subject treats a conscious urge to flex as a "go" or "decide" signal, that urge has no place in the causal process that issues in a flexing because "a neural event that precedes and correlates with the action comes before" the conscious urge (Roediger et al., this volume). But the inference here has the same form as the surprising inference about *conscious intention* that I discussed earlier. Possibly, it will be claimed that by the time the conscious urge emerges it is too late for the subject to refrain from acting on it (something that Libet denies) and that is why the conscious urge should not be seen as part of the process at issue, even if subjects think they are treating the urge as a "go" or "decide" signal. One way to get evidence about this is to conduct an experiment in which subjects are instructed to flex at time *t unless* they hear a "stop" signal (for more on stop-signal experiments, see Roediger et al., this volume). By varying the interval between the stop signal and the mean time of the completion of a full flex when there is no stop signal, experimenters can try to ascertain when subjects reach the point of no return.[13] Perhaps it will be discovered that that point is reached significantly later than time *W*. (Of course, some researchers and theorists worry about how seriously subjects' reports of their first awareness of a proximal urge or intention to flex—time *W*—should be taken.)

Libet offers two kinds of evidence to support his claim that subjects have time to veto proximal conscious urges to flex. I have already mentioned one kind: Subjects say they did this. The other kind is generated by an experiment in which subjects are instructed to prepare to flex at a prearranged time (as indicated by a revolving spot on a clock face) and "to veto the developing intention/preparation to act…about 100 to 200 ms before the prearranged clock time" (Libet 1985, p. 538). Subjects receive both instructions at the same time. Libet writes,

> a ramplike pre-event potential was still recorded…resembl[ing] the RP of self-initiated acts when preplanning is present.…The form of the "veto" RP differed (in most but not all cases) from those "preset" RPs that were followed by actual movements [in another experiment]; the main negative potential tended to alter in direction (flattening or reversing) at about 150–250 ms before the preset time.…This difference suggests that the conscious veto interfered with the final development of RP processes leading to action.…The preparatory cerebral processes associated with an RP can and do develop even when intended motor action is vetoed at approximately the time that conscious intention would normally appear before a voluntary act. (1985, p. 538)[14]

Ironically, this study indicates that a kind of RP that Libet takes to indicate the presence of an intention to flex is not actually associated with such an intention. Keep in mind that the subjects were instructed in advance *not* to flex, but to prepare to do so at the prearranged time and to "veto" this. The subjects intentionally complied with the request. They intended from the beginning *not* to flex at the appointed time. So what is indicated by the RP? Presumably, not the acquisition or presence of an *intention* to flex; for then, at one and the same time, the subjects would have both an intention to flex at the prearranged time and an intention not to flex at that time. And how can a normal person simultaneously be settled on flexing at *t* and settled on not flexing at *t*?[15] In short, it is very plausible that Libet is mistaken in describing what is vetoed in this experiment as "*intended* motor action" (p. 538; my emphasis).

If the RP in the veto scenario is not associated with an intention to flex at the appointed time, with what might it be associated? Perhaps a subject's wanting to comply with the instructions—including the instruction to prepare to flex at the appointed time—together with the recognition that the time is approaching produces a growing urge to (prepare to) flex soon, a pretty reliable causal contributor to such an urge, or the motor preparedness typically associated with such an urge. Things of these kinds are potential causal contributors to the acquisition of proximal intentions to flex in other circumstances. A related possibility is suggested by the observation that "the pattern of brain activity associated with imagining making a movement is very similar to the pattern of activity associated with preparing to make a movement" (Spence & Frith 1999, p. 27; also see Ehrsson et al. 2003). The instructions given in the veto experiment would naturally elicit imagining flexing very soon, an event of a kind suitable, in the circumstances, for making a causal contribution to the emergence of a proximal urge to flex. Finally, the "flattening or reversing" of the RP "at about

150–250 ms before the preset time" might indicate a consequence of the subjects' "vetoing" their preparation.

The experiment does not show that people have time to veto conscious proximal urges to flex, unless these subjects actually have such urges. But it does give us reason to be cautious about what is indicated by similar RPs (until "about 150–250 ms before the preset time") in other experiments. We should also be cautious about what is represented by the shorter, type II RPs produced in Libet's main experiment. Possibly, the first half of so of those RPs represents only a potential cause of a proximal intention to flex. Again, Libet's data do not show that subjects have proximal *intentions* to flex before they think they do.

Suppose that studies were to show that, contrary to what Libet claims, by time *W* the point of no return has been reached. Libet would find this result distressing; for, in his view, free will comes into play in his experiment after a subject becomes aware of a proximal urge or intention to flex. He claims that subjects can freely veto these urges or intentions. But if the point of no return has already been reached, he is wrong.

Would this imagined discovery about the point of no return actually be bad news for free will? Sketching some further background will enable me to offer an answer. Sometimes agents are indifferent between or among their leading options. Buridan's ass was in that situation regarding two equally attractive equidistant bales of hay, and scenarios of the kind at issue are sometimes called "Buridan scenarios."[16] (The ass starved to death because he had no reason to prefer either bale to the other and he was not equipped to act in the absence of such a reason.) Al, whose shopping list includes a half-pound jar of Pop's peanuts, is in a position of this kind regarding the nearest jars in the Pop's peanuts array he is facing in the supermarket. And subjects in Libet's main experiment are in this position regarding various moments to begin flexing. One difference between the latter two cases is that Al has been in many Buridan situations in supermarkets, whereas Libet's subjects, at least for a time, are unaccustomed to the task of picking a moment to begin flexing from an array of moments. A related difference is that, owing to the experimental design, Libet's subjects' moment-picking task is salient for them—after all, they have to report on when they first became aware of an urge or intention to flex—whereas Al's peanut-picking task is far from salient for him. Now, Al's taking jar *x* from the array would seem to have at least part of "its origin in preceding unconscious processes" (Libet 1999, p. 52). Although he remembers that a half-pound jar of Pop's peanuts is on his shopping list and is conscious of an array of peanuts and of grabbing a jar and putting it in his shopping cart, his taking the particular jar he takes—jar *x*—is not explained by any conscious decision on his part to take that jar nor by any conscious preference for that jar. People are well served by automatic, unconscious tie-breaking mechanisms in familiar Buridan situations.

If Al were given the following instructions and chose to cooperate, his situation would be more similar to that of Libet's subjects: "Don't grab a jar of peanuts without thinking. Instead, before you grab one, 'note…the time of appearance of [your] conscious awareness of wanting' to pick a particular jar, and be prepared to report it after you pick a jar. Use the Libet wristwatch we just gave you to identify when this mental event happens. Try to glance back and forth between the jars and the watch."[17] It might occur to Al that should he become aware of wanting to grab jar x, he would have the option of acting on that want and the option of not acting on it. If he sees himself as having no reason to prefer either option to the other, he can pull a coin out of his pocket, arbitrarily assign heads to the former option and tails to the latter, and let what he does be settled by a coin toss—or he can save a little time and effort by arbitrarily picking one of the options. Al's arbitrarily picking an option in this scenario can be described as "deciding on" that option. Someone who claims that Al is exercising free will in making his decision should point out that this is an instance of what has been called *the liberty of indifference*.[18] Liberty or freedom of this kind is nothing to write home about. If *free will* were to come into play only when agents consciously select among options in Buridan situations, it would not be much more interesting than coin tosses that are used to break ties in such situations.

It is also noteworthy that whereas Al's decision is about whether to veto or act on a conscious want and whereas Libet claims that his subjects may decide whether or not to veto a conscious urge or intention, many of our decisions have a subject matter of another kind. Al's feeling an urge to take jar x puts him in a position to make a decision that he would have been in no position to make otherwise. Just as I cannot make a decision about whether or not to accept an offer of which I am not aware (for example, an offer to buy my house), Al cannot make a decision about whether or not to act on an urge of which he is not aware. Given that he cannot make a decision about this, he cannot make a *free* decision about it. Awareness of particular options is important for deciding freely *among those options.* But this is not to say that one cannot freely decide to A unless one first acquires an urge (or intention) to A—or an urge (or intention) not to A—and becomes aware of that urge (or intention). The awareness that I have highlighted as important is awareness of options to decide among, not awareness of urges or intentions to pursue those options—urges or intentions that one can then decide to veto or decide to act on. When, some years ago, I was deliberating about whether to accept an offer of the faculty position I now occupy, I weighed the pros and cons and made a decision. I was well aware of my options. What did I decide? Did I decide to give the green light to my conscious intention to accept the job? No. At a time at which I am aware that I intend to A, I cannot proceed to decide to A—that is, to form an intention to A. I cannot form an intention that I am aware I already have. Did I decide to veto my conscious intention to reject the offer? No. I never had such an intention. Nor, as far as I can tell, did I decide to give the green light to a conscious urge to accept the offer or decide to veto a conscious urge to reject the offer. Rather, what I decided to do was to accept the job offer. The decision I made was about that, not about conscious urges.

To the extent that Libet is studying free will, he is studying it in the sphere of proximal decision making in Buridan situations or situations of a very similar kind. Generalizing from results obtained in this domain to a view about distal decisions made about important issues in situations of a very different kind would be extremely bold, to say the least. Within the sphere of the liberty of indifference, one sees oneself as having no reason for deciding to *A* rather than to *B* and vice versa. Given that fact, and given that decisions are caused, it is difficult to see why it should be thought that our not being conscious of the relatively proximal causes of our decisions in this sphere is interesting or important. Someone who assents to the following three propositions will conclude that we never decide freely: (1) the only possible location for free decisions is in the sphere of proximal decisions made in Buridan situations; (2) in such situations, both proximal decisions to *A* and proximal decisions to veto these decisions have their "origin[s] in preceding unconscious processes" (Libet 1999, p. 52); (3) no decision that has its origin in preceding unconscious processes is free. Libet rejects proposition 2.[19] But whether that proposition is true or false is not of great importance if proposition 1 is false. And Libet has given us no reason to believe that 1 is true. So even if it were shown that, by time *W*, the point of no return has been reached, believers in free human action should not lose heart. (This is not the place to defend a position on the range of situations in which free decisions and other free actions are possible. On this issue, see Mele 2006.)

A more interesting place to look for free decisions is in the sphere of distal decisions made about important practical or moral matters. This is perhaps a delicate way of expressing part of what Bandura (this volume) may have in mind when he remarks that "framing the issue of conscious cognitive regulation in terms of direct control over the neurophysiological mechanics of action production spawns unenlightening debates at the wrong level of control." Although framing the issue in terms of direct control over whether or not we veto conscious proximal urges may spawn some enlightening debates, it leaves a great deal that is relevant outside the frame.

THE ELUSIVE ILLUSION OF CONSCIOUS WILL AND THE MAGICAL SELF

I turn from Libet to Wegner. The present section's task, as I announced in my introduction, is to look into the light that Wegner's contribution to this volume sheds on "the illusion of conscious will" and on free will.

Recall my claim that if one sets the bar for free will ridiculously high, the thesis that people sometimes act freely should strike one as ridiculous. The claim is generalizable: If one sets the bar for the existence or occurrence of *anything* ridiculously high, the assertion that it exists or occurs should strike one as ridiculous. In Wegner's view, the illusion of conscious will is intimately related to a notion of "the self." He tells us (this volume) that "each self is magic in its own mind" and that "the magic self stands squarely in the way of the scientific understanding of the psychological, neural, and social origins of our behavior and thought." "Seeing one's own causal influence as supernatural is part of being human," Wegner asserts; and, apparently, this vision is part of believing that one is (or has?) a self. If the bar for the existence or efficacy of "conscious will" is set so high that we have to be supernatural beings for conscious will to exist or be efficacious in us, then, of course, conscious will should be lumped together with ghosts, fairies, and the like.

Wegner writes (this volume),

> Experience of apparent mental causation renders the self magical because it does not draw on all the evidence. We don't have access to the myriad neural, cognitive, dispositional, biological, or social causes that have contributed to the action—nor do we have access to the similar array of causes that underlie the production of the thoughts we have about the action. Instead, we look at the two items our magic selves render visible to us—our conscious thought and our conscious perception of our act—and believe that these are magically connected by our will. In making this link, we take a mental leap over the demonstrable power of the unconscious to guide action . . . and conclude that the conscious mind is the sole player.

Obviously, even people who believe that some of their conscious intentions (or the neural correlates thereof) play a role in causing some of their behavior should not believe that "the conscious mind is the sole player." After all, among the things that play a role in causing our intentions are events in the external world. And if, for example, conscious proximal intentions play a role in causing overt actions, causal processes of which we are not conscious link them to bodily motions.

So let us shelve "The Great Selfini" (Wegner, this volume) and the magical idea that the conscious mind or self is not itself causally influenced by anything and is a direct and complete cause of some of our actions. And let us turn to a more realistic hypothesis: Conscious intentions (or their neural correlates) make a causal contribution to some behavior. Wegner marshals evidence that, in some circumstances, people believe they did things that, in fact, they did not do and, in others, people believe they did not do things that they actually did. But, of course, it is a long way from these findings to the conclusion that the hypothesis just formulated is false, just as—as John Kihlstrom (this volume) observes—it is a long way from the findings about automaticity in human behavior to the conclusion that "automatic processes dominate human experience, thought, and action to the virtual exclusion of everything else." The same is true of a more precise version of the hypothesis I mentioned: Some of our conscious intentions to *A* (or their neural correlates)—for example, conscious intentions to help a stranded motorist, to start writing a certain chapter next Tuesday, to accept a job offer—make a causal contribution to some of our *A*-ings.

I can imagine a reader claiming that if my conscious intention to help a stranded motorist is itself caused by such things as my upbringing and my recently finding some change in a phone booth (which put me in a good mood), then that intention of mine (or its neural correlate) plays no causal role at all in getting me to help. Such a reader has not absorbed the point about causation that I illustrated earlier with the firecracker example. The fact that *x* has a cause does not entail that *x* is not among the causes of *y*.

Some readers who are happy enough to believe that our intentions to *A* sometimes make a causal contribution to our *A*-ings may think that if all of our decisions and intentions have causes, then we never act *freely*. Such readers should try to explain why compatibilists and causal libertarians (see the first section of this chapter) are wrong about what free action is. (Again, theorists of these kinds regard all free actions as caused, and they regard the causes of free actions as caused.) Wegner asks (this volume), "Why do we experience our actions as freely willed, arising mysteriously from the self, and why, too, do we resist attempts to explain those actions in terms of real causal sequences, events that are going on behind the curtain of our minds?" But why think of free will in terms of a magical self? Why not side with compatibilists or causal libertarians?

How radical is Wegner's position? Such passages as the following from his chapter may give the impression that it is very radical indeed:

> The magic of self…doesn't go away when you know how it works. It still feels as though you are doing things, freely willing them, no matter how much you study the mechanisms of your own behavior or gain psychological insight into how all people's behavior is caused. The illusion of self persists.

> I remain every bit as susceptible to the experience of conscious will as the next person. It feels like I'm doing things.

It looks as though part of what Wegner is asserting here is that we never do things. That assertion—interpreted literally—is radical enough to grab the attention even of a philosopher who has heard arguments for skepticism about everything under the sun. If it were true, you would not be reading this sentence, for example; instead, you would have the illusion of reading it. But it is an excellent bet that by "you" and "I" in these passages, Wegner means something like "your self" and "my self." It is true that our imaginary magical "selves" do not do anything. After all, they are only imaginary. But you and I exist, and we do lots of things. I cannot believe that Wegner would deny that.

Do we *freely* do things? That depends on how free action is to be understood. If (quoting Dennett again, this volume) "free will might be morally important without being supernatural," then maybe we sometimes act freely. If acting freely requires the existence of something that does not exist—a supernatural, magical self—then we never act freely. But I know of no good reason to understand free action in the latter way.

Exploring the extent to which the processes involved in the production of intentional actions are "automatic" or "controlled" is an important and interesting project (see, in this volume, the chapters by Bandura; Bargh; Baumeister; Kihlstrom; Miller & Atencio; Roediger et al.; Shariff et al.; Simonton; and Wegner). In my opinion (as an outsider), the project can definitely stand on its own two feet. There is no need to motivate it by importing outlandish ideas to debunk: for example, the idea that supernatural, magical selves cause intentional actions. Motivating the project in that way is rather like motivating a study of human evolution by promising to prove that human beings were not created in their present form—that is, independently of evolution—by God.

PARTING REMARKS

Work of the kind that the preceding chapters exemplify will continue to illuminate human behavior—including free human action, if we sometimes act freely. When it comes to free will in particular—that is, the power or ability to act freely—I believe that scientists and philosophers embarking on work on that topic ought to be at least a little introspective and ask themselves *why* they think free will is whatever they think it is. Once they find an answer (or find themselves stumped), they should ask themselves another question: Why do people with a different conception of free will conceive of it as they do? (Readers will have noticed a number of different conceptions of free will in this volume.) The next step would be to reflect on the relative merits of one's own conception of free will and the various alternative conceptions one encounters. One may find that some of the conceptions are self-contradictory, that others are hopelessly magical or mysterious, and that yet others suggest potentially fruitful research programs. One would expect most scientists with an experimental interest in free will to be attracted to conceptions of the third kind, and I look forward to seeing more results of research guided by such conceptions of free will.[20]

NOTES

1. Notice that the condition just offered is an alleged sufficient condition for free action, not an alleged set of individually necessary and jointly sufficient conditions.

2. As I use *consistent* (following standard philosophical practice), to say that *p* is *consistent* with *q* is to say that "*p* and *q*" is not a contradiction.

3. So if the occurrence of *x* (at time *t1*) indeterministically causes the occurrence of *y* (at *t2*), then a complete description of the condition of the universe at *t1* together with a complete statement of the laws of nature does *not* entail that *y* occurs at *t2*. There was at most a high probability that the occurrence of *x* at *t1* would cause the occurrence of *y* at *t2*.

4. For the record, as I understand *deciding* to help a stranded motorist, it is an action of forming an intention to help; and, as I see it, many intentions are acquired without being actively formed (Mele 2003, ch. 9).

5. Libertarians of the kind at issue are sometimes called *event-causal libertarians*, to distinguish them from agent-causal libertarians (mentioned by Myers, this volume). What serious agent-causationists mean by agent causation is, in my opinion, too deeply metaphysical to discuss in this chapter. For a recent critique of agent-causal libertarianism, see Mele 2006, ch. 3.

6. In this paragraph and the next two, I borrow from Mele 2006, pp. 9–10.

7. Pinker (this volume) sketches an objection to libertarianism that features chance and randomness. For fuller development of that objection—and for a reply—see my *Free Will and Luck* (Mele 2006, chs. 3 and 5). Incidentally, I am not a libertarian. I am officially agnostic about whether compatibilism is true or false (Mele 1995, 2006), and I have developed two overlapping views of free will, one for incompatibilists and another for compatibilists (Mele 2006).

8. On "psychology's working assumption of determinism," also see Myers (this volume). On various different senses of *determinism*, see Pinker (this volume).

9. Whether lay folk tend to conceive of free will in a compatibilist or an incompatibilist way is an empirical question, and there is some empirical work on it (see the chapters by Dweck and Molden and by Nichols in this volume). But it is a different question.

10. In a later article, Libet writes, "the brain has begun the specific preparatory processes for the voluntary act well before the subject is even aware of any wish or intention to act" (1992, p. 263).

11. Libet, Gleason, et al. report that "the subject was asked to note and later report the time of appearance of his conscious *awareness of 'wanting' to perform* a given self-initiated movement. The experience was also described as an 'urge' or 'intention' or 'decision' to move, though subjects usually settled for the words 'wanting' or 'urge'" (1983, p. 627).

12. At the end of the experiment, subjects can be asked how often (if ever) they were aware of proximal intentions to flex. Of course, researchers may worry about the accuracy of their reports.

13. Time *t* can be a designated point on a Libet clock, and brain activity can be measured backward from *t*. My guess is that in trials in which there is no stop signal

and in trials in which the stop signal does not inhibit a flexing, subjects will produce something resembling a type I RP. In trials in which the stop signal inhibits the onset of EMG activity, subjects might produce RPs that resemble the "veto RP" to be described shortly. In the point-of-no-return literature, that point is sometimes defined in terms of the onset of EMG activity and sometimes in terms of a completed action (using dynamometer readings as a measure of completion). A stop signal that does not inhibit the onset of EMG activity may inhibit a completed action. In the present context, the most suitable definition of the point of no return would be in terms of a completed action. Dynamometer squeezes can be substituted for flexings. Or video can be used to detect flexings.

14. For a more thorough discussion of the experiment, see Libet, Wright, and Curtis 1983 or Libet, Gleason, et al. 1983.

15. Sean Spence and Chris Frith suggest that people who suffer from anarchic hand syndrome "have conscious 'intentions to act' [that] are thwarted by . . . 'intentions' to which the patient does not experience conscious access" (1999, p. 24). Obviously, such people are not normal.

16. Ullmann-Margalit and Morgenbesser report that the example of the ass does not appear in Buridan's known writings (1977, p. 759).

17. The embedded quotation is from Libet, Gleason et al. 1983, p. 627.

18. For discussion and references, see Kane 1996, pp. 108–9.

19. Libet claims that the "decision to veto" might not "require preceding unconscious processes" and that although "factors on which the decision to veto . . . is *based*" may "develop by unconscious processes that precede the veto . . . the *conscious decision to veto* could still be made without direct specification for that decision by the preceding unconscious processes" (1999, p. 53). It is not clear what Libet has in mind here. He may be suggesting that free decisions to veto are not causally dependent on "preceding unconscious processes"; and he may be suggesting that although free decisions to veto have unconscious processes among their causes, these decisions are not deterministically caused. (Actually, he seems to be making both claims.)

20. I am grateful to John Baer and Roy Baumeister for comments on a draft of this chapter.

REFERENCES

Bargh, John, and M. Ferguson. 2000. "Beyond Behaviorism: On the Automaticity of Higher Mental Processes." *Psychological Bulletin* 126: 925–45.

Dennett, Daniel. 2003. *Freedom Evolves.* New York: Viking.

Ehrsson, H. Henrik, S. Geyer, and E. Naito. 2003. "Imagery of Voluntary Movement of Fingers, Toes, and Tongue Activates Corresponding Body-Part-Specific Motor Representations." *Journal of Neurophysiology* 90: 3304–16.

Hodgson, David. 2002. "Quantum Physics, Consciousness, and Free Will." In R. Kane, ed. *The Oxford Handbook of Free Will.* New York: Oxford University Press.

Kane, Robert. 1996. *The Significance of Free Will.* New York: Oxford University Press.

Keller, Ivonne, and H. Heckhausen. 1990. "Readiness Potentials Preceding Spontaneous Motor Acts: Voluntary vs. Involuntary Control." *Electroencephalography and Clinical Neurophysiology* 76: 351–61.

Libet, Benjamin. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8: 529–66.

Libet, Benjamin. 1992. "The Neural Time-Factor in Perception, Volition and Free Will." *Revue de Métaphysique et de Morale* 2: 255–72.

Libet, Benjamin. 1999. "Do We Have Free Will?" *Journal of Consciousness Studies* 6: 47–57.

Libet, Benjamin. 2001. "Consciousness, Free Action and the Brain." *Journal of Consciousness Studies* 8: 59–65.

Libet, Benjamin. 2004. *Mind Time.* Cambridge, Mass.: Harvard University Press.

Libet, Benjamin, C. Gleason, E. Wright, and D. Pearl. 1983. "Time of Unconscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)." *Brain* 106: 623–42.

Libet, Benjamin, E. Wright, and A. Curtis. 1983. "Preparation- or Intention-to-Act, in Relation to Pre-Event Potentials Recorded at the Vertex." *Electroencephalography and Clinical Neurophysiology* 56: 367–72.

Libet, Benjamin, E. Wright, and C. Gleason. 1982. "Readiness Potentials Preceding Unrestricted 'Spontaneous' vs. Pre-Planned Voluntary Acts." *Electroencephalography and Clinical Neurophysiology* 54: 322–35.

Marcel, Anthony. 2003. "The Sense of Agency: Awareness and Ownership of Action." In J. Roessler and N. Eilan, eds. *Agency and Self-Awareness.* Oxford: Clarendon Press.

Mele, Alfred. 1995. *Autonomous Agents: From Self-Control to Autonomy.* Oxford: Oxford University Press.

Mele, Alfred. 1997. "Strength of Motivation and Being in Control: Learning from Libet." *American Philosophical Quarterly* 34: 319–33.

Mele, Alfred. 2003. *Motivation and Agency.* Oxford: Oxford University Press.

Mele, Alfred. 2004. "The Illusion of Conscious Will and the Causation of Intentional Actions." *Philosophical Topics* 32: 193–213.

Mele, Alfred. 2005. "Dennett on Freedom." *Metaphilosophy* 36: 414–26.

Mele, Alfred. 2006. *Free Will and Luck.* New York: Oxford University Press.

Pockett, Susan, W. Banks, and S. Gallagher, eds. 2006. *Does Consciousness Cause Behavior?* Cambridge, Mass.: MIT Press.

Smith, Michael. 2003. "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion." In Sarah Stroud and C. Tappolet, eds. *Weakness of Will and Practical Irrationality.* Oxford: Clarendon Press.

Spence, Sean, and C. Frith. 1999. "Towards a Functional Anatomy of Volition." *Journal of Consciousness Studies* 6: 11–29.

Ullmann-Margalit, Edna, and S. Morgenbesser. 1977. "Picking and Choosing." *Social Research* 44: 757–85.

Wegner, Daniel. 2002. *The Illusion of Conscious Will.* Cambridge, Mass.: MIT Press.

INDEX