



Problems
from
Philosophy

Third Edition

JAMES RACHELS
STUART RACHELS

Problems from Philosophy

Problems from Philosophy

THIRD EDITION

JAMES RACHELS

Third Edition by

STUART RACHELS





Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2012, 2009, 2005 by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, any network or other electronic storage or transmission, or broadcast for distance learning. Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 1 0

ISBN: 978-0-07-353589-0

MHID: 0-07-353589-3

Vice President, Editorial: *Michael Ryan*
Director, Editorial: *Beth Mejia*
Sponsoring Editor: *Mark Georgiev*
Development Editor: *Craig Leonard*
Editorial Coordinator: *Nikki Weissman*
Marketing Director: *Allison Jones*
Marketing Manager: *Pamela Cooper*
Media Project Manager: *Bethuel Jabez*
Production Editor: *Ruth Sakata Corley*

Cover Designer: *Allister Fein*
Photo Researcher: *Natalia Peschiera*
Buyer: *Susan Culbertson*
Production Service: *Scratchgravel Publishing Services*
Composition: *11/12 Baskerville by MPS Limited, a Macmillan Company*
Printing: *45# New Era Matte by R.R. Donnelley & Sons*

Cover image: Thistle heads going to seed © IT Stock/age fotostock

Library of Congress Cataloging-in-Publication Data

Rachels, James, 1941–2003.

Problems from philosophy / James Rachels, Stuart Rachels.—3rd ed.
p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-07-353589-0

ISBN-10: 0-07-353589-3

1. Philosophy—Introductions. I. Rachels, Stuart, 1969– II. Title.
BD21.R26 2011
100–dc22

2010047489

The Internet addresses listed in the text were accurate at the time of publication. The inclusion of a website does not indicate an endorsement by the authors or McGraw-Hill, and McGraw-Hill does not guarantee the accuracy of the information presented at these sites.

About the Authors

JAMES RACHELS (1941–2003) wrote *The End of Life: Euthanasia and Morality* (1986), *Created from Animals: The Moral Implications of Darwinism* (1990), *Can Ethics Provide Answers? And Other Essays in Moral Philosophy* (1997), *The Elements of Moral Philosophy* (editions 1–4), and *The Legacy of Socrates: Essays in Moral Philosophy* (2007). His website is www.jamesrachels.org.

STUART RACHELS is Associate Professor of Philosophy at the University of Alabama. He has revised several of James Rachels' books, including *The Elements of Moral Philosophy* (sixth edition, 2010) and *The Truth about the World* (third edition, 2012), which is the companion anthology to this book. Stuart won the United States Chess Championship in 1989, at the age of 20, and today he is a Bronze Life Master at bridge. His website is www.jamesrachels.org/stuart.

Contents

Preface	x
About the Third Edition	xi
1. THE LEGACY OF SOCRATES	1
1.1. <i>Why Was Socrates Condemned?</i>	1
1.2. <i>Why Did Socrates Believe He Had to Die?</i>	3
2. GOD AND THE ORIGIN OF THE UNIVERSE	10
2.1. <i>Is It Reasonable to Believe in God?</i>	10
2.2. <i>The Argument from Design</i>	11
2.3. <i>Evolution and Intelligent Design</i>	15
2.4. <i>The First Cause Argument</i>	21
2.5. <i>The Idea That God Is a Necessary Being</i>	23
3. THE PROBLEM OF EVIL	27
3.1. <i>Why Do Good People Suffer?</i>	27
3.2. <i>God and Evil</i>	29
3.3. <i>Free Will and Moral Character</i>	33
4. DO WE SURVIVE DEATH?	38
4.1. <i>The Idea of an Immortal Soul</i>	38
4.2. <i>Is There Any Credible Evidence of an Afterlife?</i>	41
4.3. <i>Hume's Argument against Miracles</i>	49

5. THE PROBLEM OF PERSONAL IDENTITY	52
5.1. <i>The Problem</i>	52
5.2. <i>Personhood at a Time</i>	53
5.3. <i>Personhood over Time</i>	57
5.4. <i>Bodily Continuity</i>	58
5.5. <i>Memory</i>	61
6. BODY AND MIND	67
6.1. <i>Descartes and Elizabeth</i>	67
6.2. <i>Materialist Theories of the Mind</i>	71
6.3. <i>Doubts about Materialist Theories</i>	80
7. COULD A MACHINE THINK?	83
7.1. <i>Brains and Computers</i>	83
7.2. <i>An Argument That Machines Could Think</i>	84
7.3. <i>The Turing Test</i>	87
7.4. <i>Why the Turing Test Fails</i>	91
8. THE CASE AGAINST FREE WILL	94
8.1. <i>Are People Responsible for What They Do?</i>	94
8.2. <i>Determinism</i>	96
8.3. <i>Psychology</i>	99
8.4. <i>Genes and Behavior</i>	104
9. THE DEBATE OVER FREE WILL	109
9.1. <i>The Determinist Argument</i>	109
9.2. <i>The Libertarian Response</i>	110
9.3. <i>The Compatibilist Response</i>	115
9.4. <i>Ethics and Free Will</i>	118

10. OUR KNOWLEDGE OF THE WORLD AROUND US	125
10.1. <i>Vats and Demons</i>	125
10.2. <i>Idealism</i>	127
10.3. <i>What Evidence for These Views Might Be Like</i>	129
10.4. <i>Descartes' Theological Response</i>	130
10.5. <i>Direct vs. Indirect Realism</i>	133
10.6. <i>Vision and the Brain</i>	134
10.7. <i>Conclusion</i>	137
11. ETHICS AND OBJECTIVITY	139
11.1. <i>Thrasymachus's Challenge</i>	139
11.2. <i>Is Ethics Just a Matter of Social Conventions?</i>	140
11.3. <i>Ethics and Science</i>	144
11.4. <i>The Importance of Human Interests</i>	151
12. WHY SHOULD WE BE MORAL?	153
12.1. <i>The Ring of Gyges</i>	153
12.2. <i>Ethics and Religion</i>	154
12.3. <i>The Social Contract</i>	157
12.4. <i>Morality and Benevolence</i>	159
13. THE MEANING OF LIFE	166
13.1. <i>The Problem of the Point of View</i>	166
13.2. <i>Happiness</i>	167
13.3. <i>Death</i>	169
13.4. <i>Religion and the Indifferent Universe</i>	171
13.5. <i>The Meaning of Particular Lives</i>	173
Appendix: How to Evaluate Arguments	176
Notes on Sources	185
Index	196

Preface

This book is an introduction to some of the main problems of philosophy—the existence of God, the nature of the mind, human freedom, the limits of knowledge, and the truth about ethics. It begins by reflecting on the life of the first great philosopher, Socrates. Then it takes up the question of whether God exists. That question is discussed early, because how we answer it will affect how we answer almost every other philosophical question. Next comes a discussion of death and the soul, which leads us to consider some modern ideas about persons. The later chapters of the book are about whether objective knowledge is possible in science and ethics. All of these chapters are self-contained and may be read independently of one another.

Unfortunately, philosophical beliefs are often taken to be mere matters of opinion. After all, it is said, no one can prove whether God exists or whether life has a meaning. But these are topics for rational investigation. Like every responsible inquiry, philosophy is first and last an exercise in reason. We should try to figure out which beliefs are most reasonable, even if the questions are so complex that we cannot expect everyone to agree on the answers.

Some philosophers believe that philosophy is a “pure” subject that can be pursued in isolation from the sciences. I do not share that belief. Problems from philosophy are best approached by using every available resource. The philosopher W. V. Quine once remarked that “the universe is not the university.” The division of human inquiry into discrete disciplines may be useful in organizing academic departments, but it has no place in the study of philosophy. In this book you will find references to biology, psychology, history, and even to the discoveries of the magician the Amazing Randi. They are all part of a single project—the attempt to understand the world and our place in it.

About the Third Edition

Professor Heather Elliott and I read the second edition of this book aloud to one another, pausing to discuss each paragraph. Our discussions helped me to clarify arguments, fix organizational problems, and remove about 6,000 words from the text without loss of content. In addition, I have made scores of small corrections, additions, and updates. To note a few changes:

- In chapter 4, “Do We Survive Death?” I updated the scientific information about near-death experiences and removed the objection to reincarnation that concerned the size of the human population (section 4.2).
- In chapter 6, “Body and Mind,” I added a second criticism of type-type identity theory, I made it clear that the Mind-Brain Identity Theory and Functionalism are compatible (section 6.2), and I now describe Frank Jackson’s thought experiment about Mary (section 6.3).
- In chapter 7, “Could a Machine Think?” I improved the discussion of “The Objection That Computers Can Do Only What They Are Programmed to Do” (section 7.2).
- In chapter 9, “The Debate over Free Will,” I corrected the characterization of Libertarianism (section 9.2), I revised the criticism of Compatibilism (section 9.3), and I now call credit-eliminating conditions “praise-busters” (section 9.4).
- In chapter 10, “Our Knowledge of the World around Us,” I now mention Descartes’ worry that we might be dreaming (section 10.1), and I revised the Natural Theory, which is now referred to as just “the common-sense view of perception” (section 10.7).
- In chapter 13, “The Meaning of Life,” I no longer claim that money is irrelevant to happiness (section 13.2).

For their help, I thank Torin Alter, Kevin Dybvig, Heather Elliott, Carol Rachels, Alex Sherzer, Mark Walton, Chase Wrenn, and several anonymous reviewers. Special thanks go to my research assistant, Daniel Hollingshead, who did a splendid job.

—Stuart Rachels

The Legacy of Socrates

[W]hy should we pay so much attention to what “most people” think?

—SOCRATES, IN PLATO’S *CRITO* (CA. 390 B.C.)

1.1. Why Was Socrates Condemned?

Socrates was the first great philosopher in the Western tradition. The British Museum contains a statue of him that could have been carved as early as 330 B.C., 69 years after his death. The statue portrays Socrates as short and muscular, with a balding head, a beard, and a broad, flat nose—which is how his student Plato describes him.

Socrates wrote nothing, and so we depend on Plato’s books for almost everything we know about him. Plato’s books are called “dialogues” because they are written as conversations—conversations that always include Socrates as the main speaker. In Plato’s dialogues, we see Socrates pursuing the great philosophical questions in the public places of Athens. He does this not by giving speeches or lectures, but by asking a series of questions to his companions. Today, when teachers use the question-and-answer strategy in teaching a class, we say they’re employing “the Socratic method.”

In Plato’s writings, Socrates comes across as humble, sincere, and wise. Yet, as Plato tells us, Socrates was put on trial for corrupting his followers, and then the people of Athens sentenced him to death. Why this happened is something of a mystery. The Athenians were democrats, proud of their intellectual accomplishments and freedoms. Why would they put a philosopher to death for his teachings?

According to Plato, Socrates was charged with “corrupting the youth” and “impiety toward the gods.” The first charge is

vague, and no details are given. Socrates does say that his followers have used his methods on their neighbors, to the neighbors' annoyance. However, that's hardly grounds for a criminal trial. The second charge—impiety toward the gods—is also perplexing. Socrates was not antireligious, and at his trial he protested that he does believe in the supernatural. Admittedly, Socrates did hold some unorthodox views. Merely holding such views would not have gotten him arrested, but “pushing them on the streets of Athens,” as one scholar put it, might have gotten him into trouble. Still, Plato's account leaves us wondering whether this is the whole story.

So why was Socrates condemned? Although we revere him, Socrates was not popular in his own time. Socrates suggested that the charges were brought against him because people didn't like him. A great prophet had said he was the wisest of men, and Socrates accepted the mystic's praise, but with a twist. He said he was wise because he realized how ignorant he was. This may sound agreeably modest, but Socrates took it as his “divine mission” to show others that they were also ignorant. A typical conversation would have Socrates explaining to his listeners why their opinions were all wrong. Despite his gentle approach, Socrates must have been irritating.

Politics may also have played a part. The Athenians were proud of their democratic institutions, but Socrates did not share their pride. He was, according to Plato, democracy's severest critic. He said that democracy elevates men to positions of authority not because of their wisdom or their fitness to govern, but because of their ability to sway the masses with empty rhetoric. In a democracy, it is not truth that matters; it is public relations. The most influential teachers of the day were the Sophists, who taught the art of persuasion and were openly skeptical of “truth.” Had the Sophists lived 2400 years later, they would have been media consultants, radio personalities, and lobbyists.

If the Athenian democracy had been secure, then Socrates's hostility could have been taken in stride, much as Western democracies today tolerate criticism. But their democracy was not stable; it had suffered a series of traumatic disruptions. The latest had occurred only five years before Socrates's trial, when a group known as the Thirty Tyrants—led by a former student of Socrates—had staged a bloody coup and briefly held power.

Socrates denounced the Thirty Tyrants at his trial, calling them “wicked.” Nevertheless, it is easy to imagine that the leaders of Athens might have breathed easier with him gone.

At his trial, Socrates was sentenced to die by drinking a poison called hemlock. The penalty was clearly excessive, but to some extent Socrates brought it on himself. After he’d been found guilty, the rules of the court allowed him to propose his own punishment. Instead of suggesting something realistic, he proposed that he be given a lifetime pension for his services to the state—the “services” being the same activities he had just been convicted of. Only as an afterthought did he offer to pay a fine. Little wonder that the jurors accepted the punishment proposed by his accusers.

The sentence was not really all that harsh, however, because no one expected Socrates to die. Exile was an unofficial alternative. Several cities were willing to take Socrates, and his friends prepared his escape. Nobody, according to Plato, would have prevented Socrates from going. His enemies wanted him to leave, and his friends were eager to whisk him to safety.

But Socrates did not leave. Instead, he began to consider the reasons for and against escaping. He had always maintained that our conduct should be guided by reason: In any situation, he said, we ought to do what there are the best reasons for doing. Here, then, was the ultimate test of his commitment to that idea. While the chariot waited to take him away, he told Crito that if the best arguments were for going, he would go; but if the best arguments were for staying, he would stay. Then, after looking at the matter from all sides, Socrates concluded that he could not justify disobeying the court’s order. So he stayed, drank the hemlock, and died. Perhaps he had some inkling that this would secure his place in history. He warned the Athenians that it would be their reputation, and not his, that would be tarnished by his death.

1.2. Why Did Socrates Believe He Had to Die?

Socrates was not “the first philosopher”—that title is reserved for Thales, who lived a century earlier. (Why Thales? Because Aristotle listed him first.) Nevertheless, historians customarily refer to Thales and the others who came before Socrates as “pre-Socratics,” as though they inhabited some kind of philosophical prehistory and Socrates marked the real beginning.

It was not so much his doctrines as his method that set Socrates apart. He relied on arguments, insisting that truth is discovered only through the use of reason. Even the most abstract questions, Socrates believed, are susceptible to rational analysis. What is justice? Is the soul immortal? Can it ever be right to harm someone? Is it possible to know what is right and yet not do it? He believed that such questions are not merely matters of opinion. They have real answers that we can discover if we think hard enough. This idea annoyed Socrates's accusers, who, Plato says, distrusted reason and preferred to rely on popular opinion, custom, and religious authority.

Socrates believed that there were arguments compelling him to remain in Athens and face death. Could this possibly be right? What arguments could be that powerful? The pivotal issue, he told Crito, was whether he was obliged to obey the laws of Athens. The laws had commanded him. Must he not obey? Their discussion was the first recorded philosophical inquiry into the nature of legal obligation. Socrates offered three arguments for why he should accept his fate.

The Argument about Destroying the State. The first argument was that if he disobeyed the law, he would be destroying the state. Socrates explained that the state cannot exist if people do not obey its laws:

Suppose that while we were preparing to run away from here (or however one should describe it) the Laws and Constitution of Athens were to come and confront us and ask this question: "Now, Socrates, what are you proposing to do? Can you deny that by this act which you are contemplating you intend, so far as you have the power, to destroy us, the Laws, and the whole state as well? Do you imagine that a city can continue to exist and not be turned upside down, if the legal judgements which are pronounced in it have no force but are nullified and destroyed by private persons?"

If the laws of a city are meaningless, then, indeed, the city cannot survive. Socrates concluded from this that he should follow the laws of his city and drink the hemlock. More than 2400 years after Socrates's death, people still marvel at the fact that he was willing to die for such reasons.

Is the Argument about Destroying the State sound? One obvious problem is that Socrates's act of disobedience would not

have destroyed the state. If he had gone into exile, Athens would not have been “turned upside down.” Rather, the city would have gone on pretty much as before. Of course, chaos would ensue if people habitually disregarded the law. But if people disobey only once in a while, in extreme circumstances, then the state will be unharmed.

Nonetheless, one might feel that the argument makes a legitimate point. The considerations Socrates raises do prove something, even if they do not prove what he thought. They show that we have an extensive, *but not unlimited*, obligation to obey the law. We can recast the argument in this form:

- (1) It would be disastrous if the state didn’t exist.
- (2) Therefore, we shouldn’t do things that would destroy the state.
- (3) It would destroy the state if we did not in general obey the law (allowing only rare exceptions).
- (4) Therefore, we should in general obey the law (allowing only rare exceptions).

Although Socrates’s original reasoning is weak, this modified argument is much better. However, it does not follow from this argument that Socrates must stay and drink the hemlock. After all, this could be a “rare exception” that calls for disobedience.

The Analogy between the State and One’s Parents. Socrates’s second argument involves comparing our relationship with the state to our relationship with our parents. Our parents made our lives possible—they brought us into being, raised us, and educated us. We are therefore obligated to respect and obey them. The state also makes our lives possible, by providing a social environment that nurtures and sustains us. So we have a similar obligation to respect and obey the state. Again, Socrates imagines that “the Laws and Constitution of Athens” are speaking:

“Then since you have been born and brought up and educated, can you deny, in the first place, that you were our child and servant, both you and your ancestors? . . . Are you so wise as to have forgotten that compared with your mother and father and all the rest of your ancestors your country is something far more precious, more

venerable, more sacred, and held in greater honour both among gods and among all reasonable men? Do you not realize that you are even more bound to respect and placate the anger of your country than your father's anger?"

And so Socrates concludes that he must abide by the court's verdict and drink the hemlock.

This is an example of an argument by analogy. Is it sound? The problem is that our obligation to obey the law cannot be like our obligation to obey our parents because, as adults, we are *not* obligated to obey our parents. We must obey our parents when we are young because we lack judgment. As we mature, however, we learn to think for ourselves, and our relationship to them changes. As adults, we may continue to owe our parents gratitude and respect, but we no longer owe them our obedience. An adult who still "obeys his parents" as if he were 12 years old is a sad case. So this argument is weak.

Thus far, we have found no good reason why Socrates had to drink the hemlock. Neither the Argument about Destroying the State nor the Analogy between the State and One's Parents compels that conclusion. There is, however, one other argument to consider.

The Argument from the Social Contract. Socrates's third argument is the most important. As he muses about the law, he brings up the idea of a social contract—an idea that would later be developed by Thomas Hobbes (1588–1679) and Jean-Jacques Rousseau (1712–1778). The idea is that society rests on an agreement that we make with one another. Social living is a cooperative arrangement in which each of us gains enormous benefits, and in return we agree to support the institutions and practices that make those benefits possible.

What, exactly, are the benefits of social living? If we work together to maintain a stable social order, we can have education, arts, business, agriculture, entertainment, medicine, and much more. We can live in peace, have friends, and go to football games and concerts. The benefits are endless. But these good things cannot exist unless people cooperate in upholding the system that supports them. Without the system, lawlessness will prevail, and, as Hobbes put it, life will be "solitary, poor, nasty, brutish, and short."

In appealing to the social contract, Socrates again assumes the viewpoint of the laws:

“[A]ny Athenian, on attaining to manhood and seeing for himself the political organization of the state and us its Laws, is permitted, if he is not satisfied with us, to take his property and go away wherever he likes. . . . On the other hand, if [he] stands his ground when he can see how we administer justice and the rest of our public organization, we hold that by [staying in Athens] he has in fact undertaken to do anything that we tell him. . . .

“It is a fact, then,” [the Laws and Constitution] would say, “that you are breaking covenants and undertakings made with us, although you made them under no compulsion or misunderstanding, and were not compelled to decide in a limited time; you had seventy years in which you could have left the country, if you were not satisfied with us or felt that the agreements were unfair. . . . And now, after all this, are you not going to stand by your agreement?”

So Socrates concludes that he must stand by his agreement, even if it means his own death.

The social contract is the most influential nonreligious account of legal obligation. Nonetheless, critics have lodged a number of complaints against it. The most common charge is that the “contract” is fictitious. Few of us enter into the social arrangement by means of an agreement or explicit promise. Immigrants, who pledge to uphold the law when they are granted citizenship, are the exception. The rest of us are simply born into the system. Since we didn’t ask to be part of the system, one might wonder why we should care so much about this “agreement.” As one critic said, the social contract “isn’t worth the paper it’s not written on.”

What can be said in reply? To defend the idea of a social contract, we need the idea of an *implicit* promise—a promise that is not uttered but is nevertheless implied by our conduct. Socrates’s argument appeals to this idea. We undertake the obligations of citizenship, he says, not by making a pledge, but by willingly accepting its benefits. If we want to reject the contract, then we have to forego the benefits of social living.

That reply, however, is not very convincing. How *could* you forego the benefits of social living? On the one hand, you could become a survivalist and never use any electricity, or tap water,

or public roads, and so on. That would be a huge burden. On the other hand, you could simply leave the country—if, of course, you could afford to do so. But even then, you'd probably have to leave behind your friends, your family, and everything else you're familiar with. Thus, it would be tremendously difficult, if not impossible, to forego the benefits of social living. And so, it doesn't seem right to say that you've agreed to follow society's rules, just because you benefit from living in the society.

There's another problem. We need to ask how extensive an obligation the argument supports. Does the appeal to a social contract support the conclusion that we must *always* obey the law, or only that we should in general obey the law? Note that contracts are never 100 percent binding—there are always circumstances under which contractors would be released from their obligations. For example, if you and I have an agreement, and you don't keep your part of the bargain, then I don't have to keep mine. So, there are exceptions to the idea that one must always fulfill one's contracts.

But once this is conceded, the question of Socrates's fate is reopened. In the movie *The Fugitive*, Harrison Ford's character, Dr. Richard Kimble, is sentenced to die for killing his wife—a crime he did not commit. But then a train crash gives Kimble the chance to flee, and he takes it. Did Richard Kimble act wrongly? If Socrates had fled, would he have been making the same mistake? It is hard to see how either choice could be condemned on contractual grounds. The state is supposed to protect its citizens from unjust threats. In these cases, however, the state is *making* the unjust threats. Socrates and Dr. Kimble don't have an obligation to the state if the state is trying to kill them. The contract is off.

So none of these arguments proves that Socrates had to drink the hemlock. But this leaves us with an awkward question: How could he have failed to see that these were poor arguments? Part of the answer may be that these issues were new and unfamiliar when Socrates discussed them. Another reason may be that Socrates's first and third arguments are not really so bad. They reveal a lot about why we should obey the law. Socrates's mistake was not distinguishing the idea that we should *usually* obey the law from the idea that we should *always* obey the law. His arguments support the former, but he mistakenly thought they supported the latter.

There is a reason why Socrates failed to see this. His attitude toward Athens and its laws was very different from how most people today feel about “the government.” Like other Greeks, he felt powerfully bound to his city; he could not imagine himself apart from it. In the *Crito*, we are told that in all his 70 years, Socrates never stepped outside of Athens, except to represent Athens in some military campaigns. He was an Athenian in the profound way that St. Paul was a Christian.

And as for death, Socrates had no fear of it. He believed that after the death of his body, his soul would go to another world, where his questions would all be answered. Many people say they believe in paradise but are reluctant to go there. Socrates was not among them. As Alcibiades said in Plato’s dialogue, the *Symposium*: “He is absolutely unique; there is no one like him, and I don’t believe there ever was.”

God and the Origin of the Universe

I cannot think that this watch exists and has no watchmaker.

—VOLTAIRE, *LES CABALES* (1772)

2.1. Is It Reasonable to Believe in God?

Most Americans believe in God. In fact, according to a recent Gallup poll, 56 percent of Americans say that religion plays a “very important” role in their lives. And when the Pew Research Center conducted polls in different countries, it found the United States to be much more religious than other developed Western nations. In France, only 14 percent say that religion is very important to them. In Great Britain, the figure is 18 percent; in Spain, 21 percent; and in Canada and Germany, 27 percent. However, Indonesians, Pakistanis, and Indians are even more religious than Americans. Thus, in terms of professed religious faith, Americans are “closer to people in developing nations than to the publics of developed nations.”

Meanwhile, the Gallup International Millennium Survey asked people in 60 countries whether they believed in God at all. Only 45 percent said they believed in a “personal” God, while another 30 percent said they believed in “some sort of spirit or life force.” The Gallup poll found that religious belief is strongest in the elderly and the uneducated and that the rate of belief is highest in West Africa, where Islam dominates. There, 99 percent believe in a personal God. In the United States, the figure is 86 percent, while Europeans “are the most agnostic.”

But we want to know more than what people believe—we want to know whether religious beliefs are *true*. Does God exist?

Did he create the universe? Does he listen to prayers? Some people might say that such questions are about faith, and so reason is irrelevant. The pronouncements of Scripture or the Church do not require confirmation by rational argument. It is tempting to let the matter rest there—some choose to believe, others do not, and that's all there is to it. But before reaching that conclusion, we should ask what evidence is available. Can good reasons be given to support belief in God? We should not say that religious belief is “merely” a matter of faith until we are sure that rational arguments cannot be found.

The problem is that God cannot be detected by ordinary means. He cannot be seen or heard or touched, and scientific instruments are useless. Some people say they can sense his presence, but others cannot. This suggests that belief in God is only a matter of inner conviction. Nonetheless, religious thinkers have offered various arguments for the existence of God.

2.2. The Argument from Design

The Argument from Design tries to infer God's existence from the nature of the world around us—the world contains wonders that are best explained by supposing that an intelligent designer created them. As we shall see, this thought can be developed in different ways.

The Wonders of Nature. The world is full of amazing things that we take for granted. Consider, for example, the human eye. It is made of parts that work together in intricate, complicated ways. The eye has an opening through which light enters, while a mechanism makes the opening larger or smaller depending on the amount of light present. The light then passes through a lens that focuses it on a sensitive surface, which in turn translates the patterns into signals that can be transmitted to the brain through the optic nerve. If any detail is changed, the whole process stops working. Imagine that there was no hole in the front of the eyeball, or no lens, or no nerve connecting it to the brain—then everything else would be pointless.

Countless other examples could be given. The plants and animals that populate the earth are all composed of parts that work together beautifully. These living things form a delicate but viable ecosystem. Plants are eaten by animals, who are eaten

by other animals, who die and rot and feed the plants. The earth itself, moreover, is exquisitely fitted to support the life on it, being just the right distance from the sun and having just the right temperature, water, and atmosphere. Considering all this, we might wonder whether it could all have arisen by chance. It looks like the work of an intelligent designer.

This line of thought has occurred to many people, but it was William Paley (1743–1805), an Anglican clergyman and teacher at Cambridge University, who developed it most memorably. Paley’s favorite example was the eye. Even forgetting about the eye’s ingenious construction, consider how it is situated in the head: For protection it is lodged in a deep, bony socket, which protects it with a layer of fat. There are lids to further protect it. Glands are constantly producing a wash to keep the eye moist, without which, once again, the whole contraption would be worthless.

But, one might ask, so what? Having observed these remarkable facts, the argument can continue in two ways.

The Best-Explanation Argument. First, we may note that the wonders of nature require some sort of explanation. How, exactly, did the various parts of the eye come to exist? One possibility is that it all happened by chance—the lens, the optic nerve, the eyelid, and all the rest just happened to spring into being simultaneously. How lucky for us! But that is hard to believe. Yet, if chance is eliminated, what remains? Intelligent design seems to be the alternative. The eye and the other wonders of nature could have been made by God. Thus we have the Best-Explanation Argument:

- (1) Either the wonders of nature occurred randomly, by chance, or they are the products of intelligent design.
- (2) Intelligent design explains the existence of these things much better than blind chance does.
- (3) Therefore, the wonders of nature are best explained as the products of intelligent design.

The Same-Evidence Argument. A different form of the argument appeals to the idea that *we have the same evidence* that the universe was designed by an intelligent creator as we have that other things, such as telescopes, were designed. To make this

point, Paley introduced one of the most famous analogies in the history of science, that of the watchmaker.

Suppose we find a watch lying on the ground. If we inspect it, we will conclude that it was designed by an intelligent being. After all, it is made of many small parts that work together to serve a purpose. In Paley's words, "[I]ts several parts . . . are so formed and adjusted as to produce motion, and that motion so regulated as to point out the hour of the day. . . . [T]he inference we think is inevitable [is] that the watch must have had a maker." Thus, from the existence of the watch, we are entitled to infer the existence of a watchmaker. But do we not have exactly the same sort of evidence that the universe was made by an intelligent designer? The universe also consists of "parts framed and put together for a purpose"—namely, the purpose of housing intelligent life. And do we not have the same sort of evidence that some things in the universe—such as the eye, with its parts magnificently aligned for vision—were made by an intelligent designer? The Same-Evidence Argument, then, goes like this:

- (1) We conclude that watches were made by intelligent designers because they have parts that work together to serve a purpose.
- (2) We have the same evidence that the universe, and some of the natural objects in it, were made by an intelligent designer: they are also composed of parts that work together to serve a purpose.
- (3) Therefore, we are entitled to conclude that the universe was made by an intelligent designer.

Hume's Objections. These are impressive arguments, but are they sound? It would be nice if they were, because they would provide rational support for an ancient and satisfying way of understanding the universe. Unfortunately, these arguments are open to some crippling objections made by David Hume (1711–1776), the greatest English-speaking philosopher in history. Hume was a skeptic about religion at a time when skepticism could not be publicly acknowledged. So, he never came right out and said that he didn't believe. Instead, in his book *Dialogues Concerning Natural Religion*, he chipped away at the foundations of belief by exposing weaknesses in various theistic

arguments. He did not allow *Dialogues Concerning Natural Religion* to be published in his lifetime; it was published after his death, in 1779.

We may begin by noting that the Argument from Design, in all its forms, tries to infer *what causes something* from information about the thing itself. In other words, we are to infer a *cause* from its *effects*. This is a common type of inference, but it is justified only when we have a specific sort of background information.

For example, suppose we are presented with an AIDS patient and asked what caused her disease. We could reply, with confidence, that she must have the human immunodeficiency virus (HIV), and this is the cause. But why are we entitled to infer that? It is because of our past experience. In the past we have seen plenty of cases in which HIV and AIDS were linked. Physicians have treated many patients with AIDS, and in each case the virus was present. Moreover, studies have identified the mechanism that connects HIV and AIDS, and other possible causes have been ruled out. We call upon this background knowledge when we are confronted with a new case of the disease. We know what generally causes AIDS, and we apply this knowledge to specific cases.

Can we infer, in the same way, that an act of divine creation caused the universe to exist? The problem is that we lack the sort of background knowledge that would permit this inference. If we had observed God creating universes many times in the past and had never seen a universe not created by him, then we would be entitled to infer that he must have made our universe. But, in fact, we have no idea what causes universes to come into existence. We are familiar with only one universe; we did not observe its cause; and that's all we know.

The case of the watch is entirely different. When we examine the watch lying on the ground, we have lots of relevant background information. We have seen watches before, and we know that they are made by watchmakers. We can visit the factories where they are produced, and we know the names of the companies that make them. That is why we can say with such confidence that a particular watch must have been made by a watchmaker. This means that the Same-Evidence Argument is fatally flawed. Where causes are concerned, we have vastly more evidence about watches than about universes. Moreover,

these observations also cast doubt on the Best-Explanation Argument. Because we have so little experience with the creation of universes, and so little experience with the origin of natural objects, it seems too ambitious to assume that there are just two possibilities: random chance and intelligent design.

But suppose we set these points aside, and we do try to infer how the world came about. If we were serious about this, what would we conclude? What conjecture would seem most reasonable? The idea that the world was made by a single all-powerful, all-good deity would not be very plausible. After all, the world is not perfect. As impressive as the human body is, it is weak and vulnerable to disease. Some people have leprosy or muscular dystrophy. If our eyes were perfect, nobody would need glasses—and some of us, of course, are blind. Taking this into account, it might be more reasonable to conjecture that the world was made by a somewhat inept or malicious world-maker, or that we were made by an apprentice world-maker who had yet to master the craft. Again, we might notice that the world contains elements that work against one another—humans struggle to survive in an environment that is often hostile to them. This might lead us to speculate that the world was designed by a committee of world-makers working at odds with one another. Of course, no one believes such things. But the point is that these conjectures would be at least as reasonable as the idea that the world was made by a perfect God, *if* we were seriously trying to infer the nature of the Creator from the nature of the Creation.

2.3. Evolution and Intelligent Design

As the nineteenth century began, Hume's critique of the Design Argument was well known, but it was not considered decisive. Instead, Paley's argument was widely accepted. In the decades to come, Paley's books, and not Hume's, would be required reading in British universities. The reason for this is clear. The hypothesis of divine creation provided a way to account for the wonders of nature. Hume criticized this hypothesis, but he had nothing to replace it with. Why should people abandon a useful way of understanding the world when they have none better? Thus, the hypothesis of divine creation retained its appeal until an alternative account was found. In 1859, Charles Darwin provided such an account with his Theory of Natural Selection.

How Natural Selection Works. Many people assume that Darwin was the first person to come up with the idea of evolution, but he was not. In the early nineteenth century, it was already known that the earth is very old and that different kinds of plants and animals have lived at different times. Many people speculated that the appearance and disappearance of all those species might be explained by evolution. But evolution was rejected by scientific thinkers because no one could imagine how one species could change into another. Instead they accepted the theory of *catastrophism*, which holds that a series of great disasters has occurred throughout history—the last, perhaps, being Noah’s flood—in which the existing species were destroyed and then replaced by God in a new act of creation. Today catastrophism may seem bizarre, but in the early nineteenth century it was the best theory available, and many scientists accepted it. Then Darwin changed everything by explaining how evolution might take place. The Theory of Natural Selection, which he set out in his book *On the Origin of Species* (1859), supplied the mechanism needed to explain how species evolve over time.

Darwin’s genius was in realizing that three well-known facts, taken together, could explain evolutionary change. First, there is the *geometrical increase of populations*. Organisms reproduce in such numbers that, if left unchecked, the members of any one species would soon overrun the earth. (Starting from a few rabbits, there would soon be millions, and shortly thereafter trillions, until we were hip-deep in rabbits.) Second, there is the *heritability of traits*. An organism’s descendants tend to resemble it—each individual inherits the characteristics of its parents. Third, there is *variation*. Although individuals resemble their parents, they are not exactly like their parents. There are random small differences between them.

Putting these three facts together, Darwin argued as follows:

- (1) Organisms tend to reproduce in such numbers that, if all survived to reproduce again, the members of any one species would overrun the earth. This does not (and could not) happen. No species can continue to multiply unchecked. Each population reaches a certain maximum size, and then its growth stops.
- (2) It follows that a high percentage of organisms must die before they are able to reproduce. Therefore,

there will be a “struggle for existence” to determine which individuals live and which die. What determines the outcome of this struggle? What determines which individuals live and which die? There are two possibilities: It could be the result of random causes, or it could be related to the differences between individuals. Sometimes it is random. That is, the reason one organism survives to reproduce while another does not will sometimes be due to causes that have nothing to do with their particular characteristics. One animal may be struck by lightning, for example, while another is not; and this may be mere luck. But sometimes the fact that one individual survives to reproduce while another does not will be due to their different characteristics. It works like this:

- There are differences (“variations”) between members of species. Darwin did not know how or why such variations first arise, but today we know it has to do with genetic mutation.
- Some of these differences will affect the organism’s relation to its environment, in ways that are helpful or harmful to its chances for survival and reproduction.
- Therefore, because of their particular characteristics, some individuals will be more likely to survive and reproduce than others.

Here are two examples of how this happens. Suppose wolves live in an environment that is growing colder. The wolves with thicker fur will be more likely to survive and reproduce than the wolves with thinner fur. The thicker fur does not appear in response to the weather—it is just a random variation. Nonetheless, it benefits the wolves in the changed environment: wolves with thinner coats will die in the cold, while wolves with thicker coats will survive to reproduce.

Or suppose that a bird like the African finch migrates to an area in which the food supply consists of nuts. Finches with thinner beaks will not be able to crack the nuts; finches with thicker beaks will. The finches with thicker beaks will thus be more likely to survive and reproduce.

- (3) When they reproduce, organisms pass on their characteristics to their descendants. Again, Darwin did not know how this happens, but he knew that it does: An organism's offspring will have most of its characteristics. Today, we know this has to do with genes.
- (4) Therefore, since the organisms with helpful characteristics survive to reproduce, passing those characteristics on to their children, the characteristics that have "survival value" tend to be more widely represented in future generations, while other characteristics tend to be eliminated from the species. Future generations of wolves and finches will, on average, have thicker fur and thicker beaks.
- (5) In this way, a species will be modified—the descendants of the original stock will come to have different characteristics than their ancestors—and, when enough of these modifications have accumulated, we call the result a new species. This is how natural selection works.

The Theory of Natural Selection gave evolution a sound basis, and soon it replaced catastrophism as the dominant account of why different species have lived at different times. It also provided an alternative to the hypothesis of intelligent design, without appealing to blind chance. Rather than explaining the wonders of nature as God's handiwork, we can account for them as the result of natural selection.

Can Natural Selection Account for Biological Complexity?

When Darwin was enrolled at Cambridge University in the late 1820s, all students were required to read Paley's *Natural Theology*. Darwin later wrote in his *Autobiography* that "I was charmed and convinced of the long line of argumentation." At that time, the young Darwin intended to become a clergyman. He abandoned this ambition after completing an around-the-world voyage on the HMS *Beagle* from 1831 to 1835. By 1838 he had formulated the Theory of Natural Selection. After discovering natural selection, Darwin was no longer charmed by Paley's reasoning. Darwin considered the Theory of Natural Selection to be a replacement for the idea that particular aspects of nature were consciously designed. "The old argument of design in nature,"

he said, “which formerly seemed to me so conclusive, fails, now that the law of natural selection has been discovered.”

Darwin's views gradually won over the scientific community, but many people have remained skeptical of evolution, or at least skeptical of the idea that evolution eliminates the need for a designer. In the 1970s and 1980s, “creation science” came into vogue in the United States. Creationists accepted the literal truth of Genesis, and they looked for principles to explain the diversity and geographic distribution of life. Activists mounted a campaign to have creation science taught in the public schools as an alternative to evolution, but they failed because creation science was so obviously inadequate. Today the campaign has moved on to a more modest claim, namely, that “intelligent design” should be taught as an alternative to evolution in explaining the origin of species. In 1996, a scientist named Michael J. Behe wrote *Darwin's Black Box: The Biochemical Challenge to Evolution*, in which he argued that some biological systems cannot be the result of natural selection alone because they are “irreducibly complex.” Intelligent design, Behe said, is a more plausible explanation of such systems. *Christianity Today* named *Darwin's Black Box* its “Book of the Year.”

Why is natural selection supposed to be inadequate? Behe's arguments are too technical in their scientific detail for us to consider. However, it is fair to say that they have not yet convinced many scientists. Typically, the proponents of “intelligent design” point out that complex organs such as the eye are constructed of numerous parts, each of which appears to be useless except when working with the others. How are we to conceive of the evolution of all these parts? Are we to imagine a primitive eye, a primitive tear duct, a primitive lid, and all the rest developing alongside one another? The Theory of Natural Selection says that complex organs are the result of small variations that “add up” to the mature organ after many generations of evolutionary change. But there's a problem with this. Even if the fully developed eye is useful to its possessor, of what use is a half-eye that still has many generations to go before it is complete? Why should a half-eye be “selected for” and preserved for further development? These problems, say the critics, are unsolvable.

But this problem is not new. Darwin himself was aware of it. To address it, he made two points. First, he emphasized that a bit of anatomy may originally be preserved by natural selection

for one purpose, but then later, that anatomy might be used for a different purpose. Nature may jury-rig a complex structure out of whatever materials happen to be around. Second, Darwin called attention to what present-day theorists call the *intensification of function*. A biological structure that originally conferred a certain benefit might later confer that same benefit to a much greater degree.

To explain the eye, Darwin said, all we have to imagine is that a nerve only slightly sensitive to light gives an organism some small advantage. Then we can understand the establishment of the first primitive eye. From that simple thing will eventually come our complex eyes.

In living bodies, variation will cause the slight alterations, generation will multiply them almost infinitely, and natural selection will pick out with unerring skill each improvement. Let this process go on for millions on millions of years; and during each year on millions of individuals of many kinds; and may we not believe that a living optical instrument might thus be formed as superior to one of glass, as the works of the Creator are to those of man?

If the eye itself can be formed in this way, then so can the tear ducts, the eyelid, the bone, and all the rest. Take the lid, for example: Imagine that a primitive eye has been established and that in some organisms a slight variation has resulted in a small fold of skin that somewhat protects it. That variation is random (arising from genetic mutation), but because having this fold of skin makes the primitive eye better, this new feature will be selected for, and further modified, in the usual way. Darwin's analysis has withstood the test of time. Today it forms the basis for scientific thought about these matters.

After Darwin, the Best-Explanation Argument was finally refuted. Hume had pointed out its logical deficiencies, but he could not supply a better way of understanding the apparent design of nature. After taking away design as an explanation, he left nothing in its place. It is no wonder, then, that in the early nineteenth century even the brightest people continued to believe in design. But Darwin did what Hume could not do: He provided a detailed alternative, giving people something different to believe. The Best-Explanation Argument had considered only two ways of explaining the wonders of nature: chance and design. After Darwin, there was a third way.

2.4. The First Cause Argument

Today we know—or at least we think we know—that our universe began in a “Big Bang” almost 14 billion years ago and the earth was formed around 4.5 billion years ago. But what caused the Big Bang? Why is there a universe at all, rather than nothing? This question requires some sort of answer, and here, once again, it may be thought that the hypothesis of divine creation provides the answer. We may conjecture that God was the “first cause” of the universe.

This thought can be developed in at least three ways.

The Idea That God Was the First Cause in the Long Chain of Causes. One line of reasoning appeals to the principle that *everything must have a cause*. My watch was made by watchmakers working with metals extracted from the earth. Where did the watchmakers and the metals come from? The watchmakers came from their parents, while geological processes explain how the metals formed. The chain of causes can be traced back further: Those parents descended from other people, who descended from still other people; the earth itself was formed from matter moving through space; and so on. If we trace everything back far enough, we eventually come to the Big Bang, which in turn must have been caused by something. But, it is said, the chain of causes must stop somewhere; we must come eventually to the First Cause of Everything. The argument goes like this:

- (1) Everything must have a cause.
- (2) The chain of causes cannot reach back indefinitely; at some point, we must come to a First Cause.
- (3) The First Cause we may call “God.”

As a statement of faith, this line of thought may be appealing. But if it is meant to prove God’s existence, it fails. The main problem is that this reasoning is self-contradictory. It begins by saying that everything must have a cause, but then it goes on to posit the existence of something, God, that has no cause. We must choose: Do we seriously believe that everything must have a cause, or not? If we seriously believe that everything must have a cause, then we must ask what caused God. On the other hand, if we believe that “the chain of causes must stop

somewhere,” why not say that it stops with the Big Bang? After all, the Big Bang is as far back as science can go, so it is as good a place to stop as any.

Also, this argument doesn’t actually tell us that God exists. Its conclusion is merely that we may call the ultimate cause of everything “God.” Even if we agree to this, we would not be agreeing that an all-powerful, benevolent deity created the universe. The word “God” might now (for all we know) be the name of an incredibly dense point of mass and energy that preceded the Big Bang. Once this point is appreciated, it becomes clear how misleading it is to use the word “God” in this way.

The Idea That God Caused the Universe “as a Whole” to Exist.

We may think of God not as another link in the chain of causes, but as the source of the entire chain itself. The “chain of causes and effects” occurs within the universe. But now we want an explanation of the whole thing—why does the universe exist at all? Science deals only with causes and effects within the universe, and so science cannot tell us why the universe itself exists. For that, we need religion.

Thus, a different form of the argument might be:

- (1) Everything that exists within the universe is part of a vast system of causes and effects.
- (2) But the universe itself requires an explanation—why does it exist?
- (3) The only plausible explanation is that God created it.
- (4) Therefore, to explain the existence of the universe, it is reasonable to believe in God.

But this line of thought has its own problems. It is like the Argument from Design in that it attempts to infer the cause of the universe from the existence of the universe itself. The universe exists—that’s for sure—but what caused it to come into being? According to step (3), it is most plausible to suppose that God created it. And we may be inclined to agree with this, because of our religious traditions. But Hume’s observations are again relevant. To infer the cause of something, we need background information. To infer the cause of a watch, for example, we need information about what kinds of things cause watches to exist. However, we do not have this kind of

information about universes. We have never seen a universe being created, and so we can draw no conclusions about what kinds of things create them.

The First Cause Argument is like a lot of philosophical arguments in that we start with a promising idea—in this case, that divine creation might explain the origin of the universe—but run into problems when we try to formulate it into an explicit chain of reasoning. We might now be tempted to give up on the original thought. But before we do, there is one more idea to consider.

2.5. The Idea That God Is a Necessary Being

Peter van Inwagen is a distinguished contemporary philosopher who converted to Christianity as an adult, after he had already done first-rate work in philosophy. Van Inwagen writes that, after he became a Christian, the world seemed very different to him. Before his conversion, he says, “I can remember having a picture of the cosmos, the physical universe, as a self-subsistent thing, something that is just *there* and requires no explanation.” But now he can no longer think of the world in that way:

I can still call the image to mind (I *think* it’s the same image), and it still represents the whole world, but it is now associated with a felt conviction . . . that it must depend on something else, something not represented by any feature of the image, and which must be . . . radically different in kind from what the image represents.

If the universe is not “self-subsistent,” then it cannot exist by itself. Rather, it must be sustained by something else. But what sort of “something else” could sustain the whole universe? The obvious candidate for this peculiar status is God. God, according to traditional religious thought, is self-sufficient. He is the cause of everything else, but he himself has no cause. He exists eternally, without a cause and without any beginning or end.

What sort of being could be “self-sufficient”? What could be the cause of everything else and yet not itself require a cause? It all sounds very mysterious. But, according to some philosophers, there is a kind of being that could have these characteristics, namely, a *necessary* being. A necessary being is a being that, by its very nature, could not fail to exist. With this

idea in mind, we can formulate one final version of the First Cause Argument:

- (1) The universe is a dependent thing. It cannot exist by itself; it can exist only if it is sustained by something else—something that is not dependent.
- (2) God, a necessary being, is the only thing that is not dependent.
- (3) Therefore, the universe is sustained by God.

Does this argument provide good reason to believe that God exists? It is certainly full of puzzling notions. It is puzzling why the universe must be dependent. Why couldn't it exist without being supported by something else? It is puzzling why the universe could be created only by something self-sufficient—why couldn't it be created by something that also depends on something else? But perhaps the most puzzling thing is the notion of a being whose existence is necessary. What sense can be made of this?

The idea of God as a “necessary being” goes back at least as far as St. Anselm (1033–1109), the English monk who is sometimes called the father of medieval scholasticism. Anselm suggested that we conceive of God as “that than which none greater can be conceived.” God, in other words, has every possible perfection: He is perfect in knowledge, in power, in goodness, and in every other way imaginable. There is no conceivable way in which he could be better. Anselm maintained, moreover, that this is true *by definition*—trying to imagine God as having an imperfection is like trying to imagine a married bachelor or a triangle with five sides. You can certainly imagine a being *similar* to God that lacks some perfection, but then you are not thinking of God. The concept of God *is* the concept of a perfect being, just as the concept of a bachelor is the concept of an unmarried man or the concept of a triangle is the concept of a three-sided figure.

But Anselm noticed that something remarkable seems to follow from this: If a being is perfect by definition, then that being must exist. After all, if it did not exist, it would not be perfect. (A perfect being that exists is certainly better than the mere idea of a perfect being.) Hence, it is impossible that God not exist, and this is what we mean by a “necessary being.” A necessary being *could not fail* to exist. You and I are not necessary beings, because if history had gone differently, then our

parents might never have met, and so we might never have existed. But God is different. He had to exist.

This line of reasoning is known as the Ontological Argument. The Ontological Argument is unlike the Argument from Design or the First Cause Argument in that those arguments frequently occur to ordinary intelligent people. Any reflective person, considering the wonders of nature and the origin of the universe, is apt to wonder whether divine creation is needed to explain them. The Ontological Argument, on the other hand, may sound like a philosopher's trick. How can the existence of anything follow from its definition?

Yet the Ontological Argument has persuaded a number of thinkers. René Descartes (1596–1650), whom we will meet again in this book, and Gottfried Wilhelm Leibniz (1646–1716), the philosopher-scientist who along with Newton discovered calculus, both believed that the Ontological Argument is sound. Others, however, have disagreed.

In Anselm's own day, a monk named Gaunilo argued that if this argument proves that God exists, it must also prove that a perfect island exists. Suppose we say that "Islandia" is the name for our concept of a perfect island. Islandia, by definition, is perfect—it cannot be improved on. It follows, then, that Islandia must exist, because if it did not exist, it would not be a perfect island. By the same method, we could prove that a perfect banana exists, or that a perfect man exists. But this, Gaunilo observed, is absurd. Therefore, the Ontological Argument cannot be sound.

Gaunilo's reasoning shows that the Ontological Argument must be mistaken, but it does not explain the nature of the mistake. That was left to Immanuel Kant (1724–1804), considered by many to be the greatest modern philosopher. Kant observed that whether a thing is perfect depends on its properties—whether an island is perfect, for example, depends on its size, its climate, its natural beauty, and so on. Existence, however, is not a property in this sense. Whether such an island exists is a matter of whether anything in the world *has* those properties. Thus, we cannot prove that the island—or anything else—exists just by stipulating that it is "by definition" perfect. The definition of "Islandia" tells us only what Islandia would be like if it existed; it cannot tell us whether there really is such a thing. Similarly, the definition of "God" tells us only what sort of being God would be if he existed. Whether he does exist is another matter.

Conclusion. The whole business of seeking “arguments” for the existence of God might be considered suspect. People rarely believe in God because of arguments. Instead, they simply accept the teachings of their culture, or they believe in God because of some urgent inner conviction. Arguments seem irrelevant.

But arguments are not irrelevant if we want to know what is reasonable to believe. A belief is reasonable only if there is good evidence for its truth. The arguments we have considered are the most impressive attempts yet made to prove the existence of God. But none of these arguments succeed. They all contain flaws, and so they must be judged failures.

The fact that these arguments fail does not mean that God doesn’t exist—it only means that these particular arguments do not prove it. There may be other arguments, yet to be discovered, that will be more successful. In the meantime, the idea that God has created the universe may continue to play an important part in the thinking of religious believers. Divine creation may be accepted as part of a satisfying worldview, even if it is not rationally necessary. Like van Inwagen, many thoughtful people may even find this way of thinking irresistible. But, for the present at least, such beliefs cannot be regarded as rationally supported. This conclusion will not surprise those religious people who, in any case, have always regarded their convictions as matters of faith, not logic.

The Problem of Evil

Misery's the river of the world.

—TOM WAITS, *BLOOD MONEY* (2002)

3.1. Why Do Good People Suffer?

Job was prosperous. He owned both land and cattle, and he loved all ten of his children. He was a good man, generous to his neighbors and a leader in the religious life of his community. This combination of riches and virtue made him the most admired person in the region. But then everything went wrong. Foreigners invaded Job's lands, killing his servants and making off with most of his cattle. A fire destroyed the rest, leaving him penniless. Shortly afterward, a storm caused a house to collapse, killing all of his children. Then Job came down with a disease that left him covered with sores. He became so disfigured that people could not recognize him.

When Job's friends came to console him, their pity soon turned to accusation. They were pious, and they thought that Job must have done something to deserve his misfortunes. Surely, they thought, God would not allow Job to suffer unless he deserved it. "Does God pervert justice?" one asked. "God will not reject a blameless man, nor take the hand of evildoers." Another told him, "Know that God exacts of you less than your guilt deserves." But Job knew himself to be guiltless. Still, he could not explain why God had abandoned him.

This story is recounted in the Book of Job, an ancient Jewish writing that is included in the Christian Bible. It is the earliest document we have in which the problem of evil is clearly posed. Of course, the existence of evil is not a "problem" if one takes a nonreligious view of the world. From a secular point of

view, it is not hard to explain why bad things happen. The world isn't designed for our benefit. Lightning causes fires, geological forces cause earthquakes, and microbes cause disease. If we happen to be in the way, it's just our bad luck. This explains part of Job's suffering, and human viciousness explains the rest. From a secular standpoint, there is no mystery here, for we do not expect the world to be fair. Disaster strikes the righteous and the wicked alike. We may not like it, but that's the way life is.

The problem arises when we think of the world as under the control of God. God, by definition, is all good, all knowing, and all powerful. An all-good being would not want bad things to happen. He would not want children to get leukemia, or floods to destroy cities, or terrorists to kill the innocent. Instead, he would want to stop such things from happening. And he could, because he's all powerful. Why, then, do terrible things happen?

Philosophers distinguish the *logical* problem of evil from the *evidentiary* problem. The logical problem is that God and evil appear to be incompatible: If God existed, then he would create a perfect world; he would tolerate no imperfections. Yet our world does contain imperfections. So, God does not exist. That argument, however, contains a questionable assumption: that God would create a perfect world. Instead, God might have a good reason to allow some evil to exist. Maybe we can figure out what that reason is, or maybe not. But merely pointing out that there *might* be a reason allows us to say that belief in God is logically consistent with belief in evil.

The evidentiary problem, however, would remain. The evidentiary problem is that, even if God and evil are logically compatible, the existence of evil is nonetheless *strong evidence* that God does not exist. Suppose we find someone's fingerprints on a murder weapon—this is strong evidence that he committed the crime, even if it is logically possible that he was framed. Pending further investigation, the fingerprints make it *likely* that he is the killer. They point in his direction. In the same way, the existence of evil might count against belief in God, even if it does not conclusively prove that God doesn't exist. The fact that there is *so much* evil in the world makes the existence of God seem even less likely.

This is the single biggest problem for religious belief. It is certainly the problem that worries religious people the

most. Thus, religious thinkers have tried for centuries to answer Job's question: Why would God let terrible things happen?

3.2. God and Evil

Many distinguished philosophers have defended orthodox religious belief. For the first 1600 years of the Christian era, almost every important philosopher accepted a theistic worldview and worked within it. Following the Enlightenment (1660–1789), however, educated people were drawn to the worldview of modern science, and they began to develop more secular ways of understanding the world. Today philosophical inquiry is generally conducted independently of religion.

In the 1960s, however, religious thought began to make a comeback in philosophical circles. Alvin Plantinga, a Christian philosopher who now teaches at the University of Notre Dame, was an early leader in this movement. Plantinga distinguishes two kinds of theistic response to the problem of evil, a *defense* and a *theodicy*. A defense is a demonstration that the existence of God is logically consistent with the existence of evil. A defense does not pretend to reveal God's actual plan for creation. But it does aim to show that theists who acknowledge the existence of evil are not guilty of a logical contradiction. A theodicy, on the other hand, is more ambitious. A theodicy attempts to "justify the ways of God to man" by explaining how evil fits into God's actual plan for the world. Plantinga does not believe we can provide a theodicy, because we do not know the mind of God. But, he says, a theodicy is not necessary. A defense is all that is needed to permit religious people to continue in their faith.

The problem with this distinction is that a defense requires too little and a theodicy requires too much. As we have seen, a defense is easy to provide—we can simply point out that God might have a reason for permitting evil, even if that reason is unknown to us. It might be all right to stop here if we were concerned only with defending religious belief from the charge of inconsistency. But we are also concerned with the broader question of what belief is reasonable. Does the existence of so much evil make it *unreasonable* to believe in God? Is the existence of so much evil strong evidence against belief in God? To solve this problem, we need more than a mere "defense," but may not

need a full-blown theodicy. Instead, we can look for something in between. We need a plausible account of why an all-powerful, perfectly good God might permit great evils, even if we cannot be sure it is God's own reason.

Such accounts have been offered. Five major ideas have been advanced to explain why God might allow evil to exist. We will consider them one at a time.

The Idea That Pain Is Necessary as Part of the Body's Warning System. "Evil" might seem to be a mysterious notion, and it may be difficult to say exactly what evil is. But it is easy to give examples of bad things. The most obvious examples involve pain. Physical pain—especially intense, prolonged pain—is among the worst things in life, and torturers who deliberately inflict such pain are despicable people. But is God among the torturers? Consider a baby born with epidermolysis bullosa, a genetic skin disease that causes blistering all over the body, so that the baby cannot be held or even lie on its back without pain. If God created the world and everything in it, then he created epidermolysis bullosa, and he left babies vulnerable to it. What justification could there be for this?

It may be pointed out that pain has a purpose: It is part of the body's warning system. Pain alerts us to danger. When you twist your arm, it hurts, telling you not to twist it any further. And later, if your arm is still sore, the soreness tells you to leave your arm alone while it heals. Thus, if you could not feel pain, you would be much worse off. In fact, there are people born insensitive to pain, and they typically lead short and tragic lives. Similar observations hold true of other unpleasant experiences, such as fear. Fear motivates you to withdraw from danger. Faced with an angry grizzly bear, it's good to be afraid. So, it may be said, God has given us pain and fear for our own good.

This argument is persuasive as far as it goes, but it does not solve our problem. The trouble is that pain and fear are imperfect mechanisms for danger avoidance. They do not look like they were devised for this purpose by a perfect God. Sometimes we need a warning, but there is no pain. Carbon monoxide poisoning can sneak up on us without warning. Eating ice cream should be painful for the obese, but it is not. At other times, people suffer terribly even though they cannot improve their situation. The pain that accompanies advanced cancer of the

throat may tell us that something is wrong, but the information does us no good, and so the victim suffers needlessly. At still other times, pain may be so great that it is debilitating. Then it may hinder us from escaping, rather than helping: A solitary hiker who breaks his leg in the wilderness may die because he cannot drag the broken limb back to civilization. These facts do not fit the hypothesis that God has created pain for our protection. Pain looks more like the product of a hit-or-miss process of evolution than like the work of a perfect designer.

Finally, even if it is a good thing that we can feel pain, this does not explain why God creates *sources* of pain such as epidermolysis bullosa. A mother who asks why God allows her baby to have this disease can hardly be answered by pointing out that pain is part of the baby's warning system. What needs explaining is why the baby's system is being attacked in such a cruel way.

The Idea That Evil Is Necessary So That We May Better Appreciate the Good. St. Augustine (A.D. 354–430) observed that if nothing bad ever happened, then we could not know and appreciate the good. The point is partly logical and partly psychological. Logically, without the concept of evil, there could be no conception of goodness, just as there could be no notion of tallness without a notion of shortness. We could not even know what goodness *is* if we did not have evil for comparison. Moreover, psychologically, if we never suffered, we would take good things for granted and enjoy them less. How could we recognize and enjoy health if there were no disease? Thus, it is foolish to wish for a world that contains only good things.

But even if this is true, it explains only why God might permit *some* evil to exist. We might indeed need a few bad things to happen once in a while, just to remind us how fortunate we are. But it does not explain why there is so *much* evil in the world. The world contains vastly more evil than is necessary for an appreciation of the good. If, say, only half the number of people died every year of cancer, that would be plenty to motivate the appreciation of health. And because we already have cancer, we don't really need epidermolysis bullosa, much less AIDS, muscular dystrophy, cerebral palsy, spina bifida, diphtheria, Ebola, heart disease, diabetes, Alzheimer's, bubonic plague, and a thousand other ailments.

The Idea That Evil Is Punishment for Wrongdoing. Job's friends believed he must have done something to merit his suffering, because a just God "will not reject a blameless man." Thus, they concluded, Job must have deserved what he got.

The idea that evil is punishment for wrongdoing goes back to the creation story in Genesis, which says that humans originally inhabited a world without evil. But the first humans, Adam and Eve, rebelled against God and as a result were cast out of paradise. The idea expressed in this story is not that you and I are still being punished for Adam and Eve's wrongdoing. That would obviously be unfair. The point is that we are all sinners, and our existence in a world of evil is somehow explained by that fact. We suffer because we bring it on ourselves.

What are we to make of this? It would make sense if there were some correlation between one's moral character and how one fares in life. If evil is punishment for sin, we should expect the worst sinners to suffer the worst disasters. But no such correlation exists. Disaster strikes the righteous and the wicked alike, without apparent regard for their virtue. Some of the best people get Parkinson's disease, while some of the worst people enjoy good health.

And what about innocent babies, who sometimes have terrible diseases and die horribly? The doctrine of original sin was introduced into Christian thought partly to deal with this problem. According to that doctrine, we are all born in sin, so even babies are sinners. But it would be absurd to conclude from this that a newborn baby deserves to have epidermolysis bullosa.

The deep problem with using the concept of sin in this way is that it separates sinfulness from what a person actually does. In the ordinary moral sense, what a person deserves depends on his or her behavior. If you deserve ill, you must have done something to deserve it. No doubt all of us have behaved badly at one time or another, so perhaps each of us deserves some degree of punishment. But we don't deserve a horrible disease unless we have done something pretty awful, and a baby hasn't done anything bad at all. Therefore, while this conception of "sin" might have religious significance, it has little to do with moral desert. As Job knew, not all calamities are deserved.

3.3. Free Will and Moral Character

The ideas we have considered so far—that pain is part of the body’s warning system, that evil is necessary for appreciating the good, and that evil is punishment for wrongdoing—are unconvincing. But two other ideas are more promising. Together, they make up the most plausible theistic response to the problem of evil.

The Idea That Evil Is the Result of Human Free Will. God could have made the world, it is said, without people in it. It could have been a beautiful place, with sunsets and waterfalls, with daffodils and butterflies, and God could have been pleased with it. But God went one step further and chose to include us as well. What difference did that make? Humans are not just another kind of animal. We are moral agents, able to choose what sorts of persons we’ll be, and so we’re responsible for our choices. We are capable of love and friendship, and we can envision and accomplish great things. Thus, without human beings the world would be a poorer place. That is why humanity was part of God’s plan—or at least why God might have chosen to create us.

But there is a catch: In order to make us moral agents, rather than mere robots, God had to endow us with free will. Freedom makes humans special. In giving us the power of free choice, however, God had to allow that we might sometimes choose badly. He could not have given us free will and at the same time fixed it so that we could never do wrong—we might then have the appearance of free will, but we would not have the real thing.

The result is that human beings have the power to choose evil, and sometimes we do. Therefore, in addition to love, kindness, sympathy, and heroism, humans also bring to the world murder, cruelty, rape, and war. But God is not responsible for that; we are. God is responsible for the overall design of the world and for creating us. His creation is good. Our contribution to it, however, is not always good.

This line of thought is called the Free Will Defense. The Free Will Defense shows that the best world God could have created might contain evil, because the best world might include creatures with free will. Compare:

- (a) A world without humans in it, which would not include any of the bad things people do, but would

also not contain any of the results of human consciousness, creativity, and virtue

- (b) A world with humans in it, which would include human wickedness but also the good things that humanity brings

If you think that (b) is better than (a), then perhaps the best world God could have created would have evil in it. This vindicates the perfect goodness of God. After all, we could not expect more of him than to create the best world possible.

Still, this cannot solve the problem completely. We need to distinguish two kinds of evil, *moral evil* and *natural evil*. Moral evil is the evil that people cause by their own actions—murder, rape, war, and so on. Natural evil, on the other hand, is the evil that human beings do not cause—disease, earthquakes, droughts, floods, and other natural disasters. Humans may be responsible for babies born addicted to narcotics, but we are not to blame for babies born with epidermolysis bullosa. Thus, even if the Free Will Defense explains why God would allow moral evil, it does nothing to explain why he would create natural evil. Therefore, it can be only part of the story.

The Idea That Evil Is Necessary for the Development of Moral Character. Suppose that, at the beginning of time, God decided to make a world that included human beings, with all the capacities of thought and action that make us who we are. This decision would restrict the kind of world God could make, because the world would have to be suitable for creatures like us. What sort of world would that be? What sort of environment would allow us to grow and develop as human beings?

Imagine a world that was “perfect” in the ordinary sense. In this world, the temperature would be a balmy 72°F year-round, and nothing bad would ever happen—there would be no destructive earthquakes, fires, or floods and no diseases. There would be no dangerous predators. Moreover, there would be no hunger. Every tree would bear abundant fruit. Everyone would live to a ripe old age in perfect health, and everyone would enjoy life. The environment would be beautiful in a way that we never grow tired of.

You might ask, If God is perfectly good, why didn’t he create that sort of world? The answer is that if we lived in *that* sort

of world, we would not be the creatures we are. In a “perfect” world there would be no problems to overcome, and thus no occasion for the development of moral character. There would be no courage, because there would be no dangers to face. There would be no helpfulness or generosity, because no one would need help. All the other virtues—such as kindness, compassion, perseverance, and creativity—would also go by the way-side, because we develop such qualities only by striving against adversity. If we lived in a perfect environment, there would be no work to be done. We would be slugs—slugs who enjoy life, but slugs nonetheless.

Thus, we have a plausible explanation for natural evil: In order to create human beings as creatures with moral character, God had to place us in an environment in which those qualities could be developed. That means an environment with problems to be solved and evils to be overcome. The world that we actually inhabit seems well suited for the development of people like us.

A Complete Account? The most plausible response to the problem of evil combines the Free Will Defense with the idea that evil is necessary for an environment in which human beings can develop and flourish. The former explains moral evil, while the latter explains natural evil. Together, they seem to give a more or less complete account of the evils we face. They explain why a perfectly good, all-powerful God might have created a world like the one we actually inhabit. We need not pretend that this account is a “theodicy” in Plantinga’s sense—perhaps God had other purposes in mind, which we are ignorant of. But this account does seem to be the best available way to square the hypothesis of God with the existence of evil.

The picture may be rounded out by adding another idea, namely, that human life as we know it is a mere prelude to the eternal life we will enjoy after we die—or rather, after the bodies we inhabit die. As Peter van Inwagen (1942–) puts it:

Every human being has an eternal future (and, therefore, the human species has an eternal future). We are now living, and have been living . . . within a temporary aberration in human history, an aberration that is a finite part of an eternal whole. When God’s plan of atonement comes to fruition, there will never again be undeserved suffering

or any other sort of evil. The “age of evil” will eventually be remembered as a sort of transient “flicker” at the very beginning of human history.

If, indeed, the evil that concerns us is only a “transient flicker” in human history, then the problem does not seem nearly so pressing.

Should we be content with this and conclude that the problem of evil has been overcome? That is different from asking whether we actually believe the theistic story. We are discussing whether the existence of evil is evidence against God’s existence. If evil can be justified within a theistic framework, then even a nonbeliever should agree that the problem of evil has been solved.

But three problems stand in the way of such a conclusion. The first has to do with moral evil. The Free Will Defense says that God wants us to be free, so he tolerates the bad effects of human misconduct. This might explain why God doesn’t step in to prevent small offenses—perhaps those things should be allowed to occur. But why wouldn’t God intervene when people are about to do something awful, like abuse an innocent child? It is not plausible to say, “Intervening would turn the attacker into a robot” or “Respecting the abuser’s free will is more important than protecting the child.” We would not accept those excuses from a police officer who passively witnessed the crime, so why should we offer those excuses on God’s behalf? The Free Will Defense explains why God doesn’t intervene all the time, but it doesn’t explain why God allowed the Holocaust or why God allows isolated cases of extreme cruelty.

The second problem has to do with the theist’s explanation of natural evil. The need to develop moral character might explain why there is *some* evil in the world, but there is far more evil than is necessary for such a purpose: There is stunning, overpowering misery that crushes the life out of people. If we already have AIDS, muscular dystrophy, cerebral palsy, and spina bifida, why do we also need malaria, tuberculosis, and Ebola? If the people of Haiti are already poor and hungry, why do they need an earthquake on top of it? Nothing in our account addresses why there is *so much* suffering caused by natural events.

Consider the true story of:

a man who drove a cement mixer truck. He came home one day for lunch; his three-year-old daughter was playing in the yard, and after lunch, when he jumped into his truck and backed out, he failed to notice that she was playing behind it; she was killed beneath the great dual wheels.

If this man were told the story about free will and character development, he might find it unconvincing, not because he is blinded by grief and guilt, but because he hasn't been told why *this* sort of thing must happen, in addition to all the other troubles that plague us. The amount of evil in the world could be reduced by two-thirds, and there would still be more than we could handle.

The third problem has to do with nonhuman animals. The traditional debate about evil has centered on human beings. But human life and history are only a small part of nature and its history. Countless animals suffered terribly in the millions of years that preceded the emergence of *Homo sapiens*, and none of the ideas we have been considering address their suffering. Animals are not sinners, they do not have "free will," they do not develop moral character, and they are not going to heaven. What about them? Charles Darwin (1809–1882) made this point forcefully:

That there is much suffering in the world no one disputes. Some have attempted to explain this in reference to man by imagining that it serves for his moral improvement. But the number of men in the world is as nothing compared with that of all other sentient beings, and these often suffer greatly without any moral improvement. A being so powerful and so full of knowledge as God who could create the universe, is to our finite minds omnipotent and omniscient, and it revolts our understanding to suppose that his benevolence is not unbounded, for what advantage can there be in the sufferings of millions of the lower animals throughout almost endless time?

We are left, then, with this conclusion. Our survey of ideas about how to reconcile God with evil has turned up various thoughts that might be useful. But none of them succeed in reconciling the existence of God with the amount of suffering in the world. All the gratuitous, pointless evil that exists not only poses a serious problem for the believer. It is also a reason why someone who is considering whether to believe in God might decide not to.

Do We Survive Death?

Souls and bodies have one birth and growth, and the same wastes of age decay them both.

—LUCRETIVS, *ON THE NATURE OF THINGS* (CA. 50 B.C.)

4.1. The Idea of an Immortal Soul

Socrates, the first great philosopher in the Western tradition, was condemned to death by a court in ancient Athens. On the day of his execution, Socrates and his friends discussed the immortality of the soul. His friends were distressed that he was about to die, but Socrates told them not to worry. After all, he reminded them,

By death do we not mean simply the departure of the soul from the body? Being dead consists, does it not, in the body having been parted from the soul and come to be by itself, and in the soul having been parted from the body and being by itself. Can death possibly be anything other than that?

Since the soul continues to exist, what is there to fear? In fact, Socrates says, the soul of a philosopher will be better off separated from his body, because the body is a hindrance to the pursuit of truth. Socrates was confident that he would live on in a disembodied state, and he was eager to see what it would be like.

Today we think of “the soul” as a religious idea, but Socrates did not. He thought it was a fact, independent of religion, that each human being is composed of a physical body and a non-physical soul. Similarly, he believed on rational grounds that the soul—the part of you that perceives, thinks, and feels—will never die.

Socrates knew that this last claim was open to doubt. As Cebes told him,

Your view about the soul is one that people find very hard to accept; they suspect that, when it has left the body, it no longer exists anywhere; on the day when a man dies, his soul is destroyed and annihilated; immediately upon its departure, its exit, it is dispersed like breath or smoke, vanishing into thin air, and thereafter not existing anywhere at all.

To counter this, Socrates gave several arguments to show that the soul cannot be destroyed. His most plausible argument appeals to the soul's simplicity.

Destroying something, Socrates says, means breaking it down into its component parts. We could destroy a shirt, for example, by tearing it to bits so that nothing remained except scraps of cloth. But nothing like that can be done to the soul. The soul is simple. It has no parts that can be detached from one another. Moreover, because the soul is not a physical thing, it cannot be burned, crushed, or mangled. Thus, it cannot be destroyed.

Socrates's friends were convinced by this argument. Unfortunately, however, the argument is not as strong as they thought. First, even if we assume that souls exist, why must we assume that they are simple? Couldn't the soul have one part for thinking, another part for perceiving, another part for feeling, and so on? Second, as Cebes suggests, ceasing to exist might not involve coming apart. Think for a moment about how a simple soul could come into existence. It couldn't be constructed from smaller parts, since it has no small parts. Rather, it would have to come into being by simply appearing one day, as if by magic: poof! Similarly, it might go out of being simply by vanishing.

Even though Socrates's arguments were flawed, his idea of an everlasting soul has had a great influence. It was adopted by Christian thinkers and became part of the Church's teachings. For this reason, many people believe that the immortal soul is a biblical doctrine. However, the Bible teaches no such thing. Instead, the Bible—or at least the New Testament—holds out the hope that, some time after we die, God will bring us back to life.

These two notions are very different. According to Socrates, you will survive death because *there is something in you that is indestructible*. Or, to put it another way, *you yourself are indestructible*.

To survive, you do not need to be resurrected; you do not need the sustaining power of God, or anything like it. Your soul is immortal by itself, in its own nature. According to St. Paul, however, there is nothing about you that is indestructible. When you die, you are well and truly dead. Your only hope is that God will restore you to life, just as Jesus was resurrected. Moreover, the Bible does not talk about people existing in a disembodied state. Rather than imagining the afterlife as a realm of pure spirit, St. Paul speaks of getting a new body.

The intellectual climate today is unlike that of Socrates's Athens. The hope of eternal life remains with us, but speculative arguments for the immortality of the soul no longer carry much weight. The reason is that modern science goes against Socrates's idea, in two ways.

First, the idea of a "soul" has disappeared almost entirely from the scientific study of human beings. In modern science, the idea that each of us is a body-plus-soul has been replaced by theories that focus on the relation of thought to the brain. Psychologists are apt to talk about behavior, the emotions, cognitive capacities, perceptual systems, genetics, the environment, natural selection, and the brain—but the "soul" does not enter the picture, except perhaps for a brief nod when scientists give popular expositions of their work. Even then, the word will be in quotation marks.

Second, a great deal of evidence suggests that consciousness is not possible unless your brain is functioning in the right way. We see this again and again: When you are hit on the head, events occur in your brain that cause you to lose consciousness; if the supply of oxygen to your brain is interrupted, you lose consciousness; if an anesthesiologist "puts you under," you lose consciousness. In each case, consciousness is restored only when your brain begins to function normally again. From this we may infer that when you die and your brain shuts down for good, you lose consciousness permanently. The lights go out, and there is no way to turn them back on.

Nevertheless, hope remains. The great religions teach that we survive bodily death, even if our best scientific evidence does not agree. Many believers are content to let the matter rest there, as a matter of faith. Others, however, want more than religious hope. They look for factual evidence to set against the cold pessimism of science. Does any such evidence exist? Many websites and TV shows say that it does. The purported evidence

is of three kinds: near-death experiences, instances of reincarnation, and mediums who communicate with the dead.

4.2. Is There Any Credible Evidence of an Afterlife?

Go into any large bookstore, and near the philosophy section you will find a shelf labeled “New Age” or “Spirituality.” There you will find books about angels, astral voyages, crystals, flying saucers, the mystic powers of pyramids, and tarot cards. Topics come and go. Books about alien abductions, Bigfoot, and the Bermuda Triangle might also be there, but those topics are no longer popular. These books are collected in one place, not because they concern a common subject, but because they appeal to a common sensibility. They see the world as having mysterious dimensions, full of significance but ignored by the mainstream culture. In their pages, scientists are viewed as a closed-minded elite who refuse to acknowledge the “research” done by psychics, mediums, and other spiritualists. While scientists ignore these books, ordinary people read them. According to a recent survey, nearly half of Americans believe in extrasensory perception, or ESP; nearly one-third of Americans believe in ghosts (and nearly one-fourth say they’ve seen a ghost or felt its presence); and nearly one in five accepts the existence of spells or witchcraft.

Near-Death Experiences. In 1975, a medical student named Raymond A. Moody, Jr., published a short book called *Life after Life*. Moody, who had previously earned a Ph.D. in philosophy, wrote that he had interviewed 50 people who had undergone remarkable experiences when they were near death. Typically, he said, they would hear themselves pronounced dead and feel themselves floating out of their physical bodies. Looking down, they would see the doctors trying to revive them, and then they would move on, perhaps passing through a tunnel toward a bright light where they would be greeted by the spirits of old friends and loved ones. A “being of light” would seem to be in charge. They would feel intense joy and peace. But then, as the doctors brought them back to life, they would be pulled back into their physical bodies. Although Moody disavowed any intention of proving the existence of an afterlife, the title of his book—*Life after Life*—proclaimed otherwise.

Life after Life was an instant best seller. In it, Moody coined the phrase “near-death experience,” which has become an established part of the English language. *Life after Life* also inspired the creation of many similar books. Some of these books report additional experiences like the ones Moody describes. Others detail further aspects of “the light,” such as its healing powers. One book even suggests that we now know enough about “the other side” to make plans for what to do when we get there—it is called *What to Do When You Are Dead: Living Better in the Afterlife*. Today, there is a whole “field” of near-death research, with its own professional association and its own journal. Moody, meanwhile, went on to become a “past-life therapist,” helping people overcome their psychological problems by facing traumas from their previous lives. He has also developed a method for communicating with the dead by gazing into mirrors. Moody explains that he discovered this last technique after “studying ancient writings as well as an archaeological site in Greece.”

What are we to make of all this? Setting aside the sillier aspects of these books, do near-death experiences provide evidence of an afterlife? There are several obvious problems. For one thing, it is hard to judge how frequent these experiences are and what they are really like. Even in the most favorable circumstances, people are bad at remembering their dreams. Moreover, psychologists know from long experience that people will say whatever they think will please the interviewer, or they will fudge what they say in order to make themselves appear more interesting.

“Investigator bias” is also a problem, even for the best researchers. It is hard to avoid emphasizing evidence that supports one’s own theory while playing down evidence that goes against it. In conducting interviews, this bias may take the form of asking leading questions, interpreting the responses to fit one’s preconceptions, and then writing summaries that leave out the material that goes against one’s preferred theory. Mainstream scientists are aware of these dangers and take precautions against them. One precaution is to let other scientists review the tapes and transcripts of their interviews. Moody, however, has never let anyone see the interviews that his book is based on, so no one knows how much investigator bias crept into his work.

This might not matter if impartial investigators could reproduce Moody’s results. After Moody’s book came out,

physicians with no axes to grind began to note the experiences of their own patients. They found that in most cases the dreams of patients undergoing medical crises have little to do with happy reunions or feelings of peace. An editorial in the *Lancet* (the official publication of the British Medical Association) commented that “Of male survivors of cardiac arrest, 80 percent had dreams of violence, death, and aggression, such as being run over by a wheelchair, violent accidents, and shooting their way out of the hospital only to be killed by a nurse.” Even when patients did dream about peace and light, the details varied with their education level. Less-educated patients dreamed of being met by friends and loved ones, while better-educated patients did not.

But even if we set these problems aside and grant that people have the experiences that Moody describes, a further question arises: What is the best explanation of those experiences? “Contact with the next world” is one hypothesis. Here are some others:

- *Activity in the brain.* Researchers can induce out-of-body experiences by stimulating an area of the brain known as the superior temporal gyrus. Also, people who achieve out-of-body experiences through prayer and meditation show reduced activity in their superior parietal lobes—the area of the brain that generates one’s sense of space and time. Perhaps those areas are involved in near-death experiences.
- *Influence of drugs.* The drug ketamine, which is sometimes used as an anesthetic, produces all the features of near-death experiences. One psychiatrist compared the reports of near-death experiences to the reports of people undergoing drug-induced hallucinations and found them virtually identical.
- *Oxygen deprivation.* Oxygen deprivation can cause hallucinations, and someone whose heart has stopped beating will lose oxygen. The “tunnel” that some people see when they’re near death might simply be a narrowing of the visual field, like the one people experience before they faint.
- *REM sleep.* We all know what it’s like to lie in bed half-awake and have strange thoughts, partly connected with the world around us and partly hallucinatory. The

neurologist Kevin Nelson believes that near-death experiences are manifestations of REM (“rapid eye movement”) sleep. On Nelson’s theory, the patient’s REM state is brought on by the stress of nearly dying; the feeling of being dead is brought on by sleep paralysis; and the tunnel and the light are due in part to the loss of blood to the retina. So, some of the mechanism involved in sleep might also be involved in near-death experiences.

With these other explanations available, there is no need to postulate anything as radical as “contact with the next world” to explain the experiences.

For these reasons, most philosophers, like most scientists, have little interest in Moody’s work. In 1988, however, philosophers were startled when one of their own reported a near-death experience. A. J. Ayer was an enormously respected philosopher who had no patience with supernatural notions. His best-known book, *Language, Truth and Logic* (1936), argued that metaphysics and religion are nonsense and that science is the only reliable guide to truth. Throughout a long and distinguished career, he never abandoned that creed.

One year before his death, Ayer was hospitalized for pneumonia. While he was in intensive care, his heart stopped beating for four minutes. Upon awakening, he said to a friend: “Do you know that I was dead? The first time I tried to cross the river I was frustrated, but my second attempt succeeded. It was most extraordinary, my thoughts became persons.” Ayer, who was later amazed that he had said such things, interpreted these statements as reflecting his classical education—the “river” was obviously the River Styx, which serves as a gateway to the afterlife in Greek mythology. But he also reported having had another experience while unconscious: “I was confronted by a red light, exceedingly bright, and also very painful, even when I turned away from it. I was aware that this light was responsible for the government of the universe.” Reflecting on this remarkable vision, Ayer concluded that the explanation was probably naturalistic: “The most probable hypothesis,” he said, “is that my brain continued to function although my heart had stopped.” Nevertheless, he added that “My recent experiences have slightly weakened my conviction that my genuine death, which is due fairly soon, will be the end of me, though I continue to hope that it will be.”

Reincarnation. According to the hypothesis of reincarnation, each of us has lived before, and each of us will live again. Perhaps you are the reincarnated spirit of a seventeenth-century Frenchman, and perhaps, someday, a twenty-third-century soldier will be the reincarnation of you.

Could this sort of thing be true? There are some obvious difficulties. One problem is that, if reincarnation occurs, why aren't babies born with adult personalities? After all, a baby is supposed to "be" someone who was 60 or 70 years old. If the baby has the old person's psyche, why doesn't the baby have the old person's knowledge and experience? The philosopher Paul Edwards comments that "It is little less than scandalous that no reincarnationist has ever attempted to reply to this argument. It is as if Christian theologians had never attempted to face the problem of evil."

Perhaps this objection could be outweighed by good empirical evidence that reincarnation occurs. Is there any such evidence? The most famous case of apparent reincarnation was Bridey Murphy. In 1952, a Colorado businessman and amateur hypnotist named Morey Bernstein placed Virginia Tighe, a Chicago housewife, into a hypnotic trance. While in the trance Mrs. Tighe began talking in an Irish accent and said that her name was Bridey Murphy and she was born in Ireland in 1798. In subsequent sessions Mrs. Tighe provided impressive details about her life as Bridey, from her marriage to a barrister in 1818 to her death on a Sunday in 1864. The newspapers got hold of the story, and soon everyone knew who Bridey Murphy was. Bernstein, who had long been an enthusiast of reincarnation, eventually wrote a book, *The Search for Bridey Murphy* (1956), that rose to the top of the best-seller lists.

While Bridey's picture of nineteenth-century Irish life enthralled ordinary people, it did not impress the experts. W. B. Ready, an Irish scholar at Stanford University, said that she depicted "an Ireland that never was, save in the minds of the uninformed and the vulgar." But because of the high level of public interest, attempts were made in Ireland to verify the details of her story. The results were unexciting. No record of Bridey's birth or death could be located, despite the existence of birth and death registries in the cities where she supposedly lived. Cork city directories are virtually complete from 1820 on, but no mention of her family, including her barrister husband,

could be found. Bridey said her husband taught law at Queens University, but no record of that could be found, either. On the other hand, it could be said that she knew the names of various places in Ireland, and she had heard of Queens University.

The public discussion of the Bridey Murphy case ended on a bizarre note when the *Chicago American*, a Hearst newspaper, published an exposé “revealing” that when Virginia Tighe was a girl, she would visit a neighbor, an Irish woman named Mrs. Bridey Murphy Corkell, who would tell her tales of the old country. After this story had been circulated and, as a result, people had lost interest in Bridey, the *American’s* story was itself exposed as a hoax.

In recent years, the leading defender of reincarnation has been Dr. Ian Stevenson (1918–2007), a psychiatrist who taught for many years at the University of Virginia Medical School. Dr. Stevenson devoted his life to collecting and investigating cases of alleged reincarnation. Many involve stories told by small children about their previous lives. These cases almost always occur in countries where belief in reincarnation is widespread, and the children almost always stop telling the stories at age 7 or 8. Sometimes the child’s family will seek out the child’s “other” family, from the previous life, to verify the child’s report. Dr. Stevenson was convinced that this research proves the reality of reincarnation.

It is hard to evaluate such evidence, but we might note the assessment of Champe Ransom, who was Stevenson’s research assistant for three years—an experience he describes as “interesting but painful.” In a report written for Dr. Stevenson, Ransom describes a number of problems with the Stevenson team’s research methods. While he did not publish or circulate the report beyond the Stevenson laboratory, he did provide a summary for Paul Edwards. Ransom’s comments read like a checklist of precautions that any scientist would think obvious:

Three notable [flaws] are that leading questions were being asked, that the questioning period for each case was quite brief . . . and that the time elapsed between the alleged occurrence of the event and the investigation was quite long, frequently years. Stevenson’s cases seldom, if ever, contain any study or discussion of the child’s story-telling inclinations to get attention. Similarly, there is

no investigation of the playmates of the subject and the extent of *their* knowledge of the events in question. Still another set of factors Stevenson neglects concerns human fallibility. There are subtle distortions of memory; there is the tendency to unintentionally “fill in” a story in order to make it complete; there is a general lack of caution of many people in reporting what they have observed; there is the general unreliability of detail of personal observation; there is witness bias; and there is witness suggestibility.

These are not esoteric matters of scientific technique. After all, science is not an exotic art; it is just the persistent use of human intelligence in the attempt to figure out the truth about the world. Ransom’s points are “scientific” because, as scientists are aware, no investigation that ignores them has much chance of finding the truth.

As for confirming the children’s memories by checking with the families from their previous lives, Ransom notes:

[I]n only 11 of the approximately 1,111 rebirth cases had there been no contact between the two families before investigation was begun. Of those 11, seven were seriously flawed in some respect. What this means is that in the great majority of cases, the two families had met years before a scientific investigation began, and that the likelihood of independent testimony was quite small.

Psychics Who Claim to Communicate with the Dead. Though many people question the quality of Stevenson’s work, nobody doubts his integrity. When we turn to the psychics, however, we enter a world of fakery and fraud. Modern spiritualism began in 1848 in Hydesville, New York, when two girls, Katie and Maggie Fox, claimed to communicate with a ghost by means of mysterious rapping sounds. The Fox sisters were a sensation, and soon a host of others were making similar claims. In later years, Katie and Maggie admitted it was just a trick—they had made the sounds by knocking or kicking anything handy. Later, spiritualists would develop more sophisticated techniques. Trumpets would float while spirits made music; ghostly images would appear on exposed film; tables would float in the air; strange breezes and odors would waft through the room; and otherworldly visitors would inhabit the bodies of psychics and scrawl messages on slates.

To professional magicians, it was plain that the spiritualists were using tricks of the trade. Harry Houdini was among those who exposed the frauds. But as the spiritualist craze spread to Europe, scientists and members of learned societies declared the psychics to be genuine. William James, the brother of Henry James and one of America's most distinguished psychologists and philosophers, wrote of one medium: "I should be willing now to stake as much money on Mrs. Piper's honesty as on that of anyone I know."

It should not be surprising that intelligent people can be taken in by fakery—the presumption that sincere-seeming people are honest is very strong, and even a clumsy magician can produce startling illusions. Why should training in psychology or physics make someone immune to the magician's art? James Randi, a magician who regards psychics as unethical practitioners of his trade, put the point well. If you want to catch a cheat, he said, give the job to a magician, not to a scientist.

The most common technique used by spiritualists today is "cold reading," a method by which the psychic appears to produce information when in fact the information is being provided by the client. As the medium "makes contact" with someone in the spirit world, she will make vague statements, and the "sitter," as the client is called, will react to them. The medium might say, "I feel like there's a G- or J-sounding name attached to this . . . does that mean anything to you?" Or, "He's pointing to his head . . . there's something about a head here . . . does that mean anything to you?" The sitter will think of something significant suggested by these references, and then the process is repeated as the medium extracts details. If the psychic is addressing a group, achieving results is even easier: Only one member of the group needs to come up with a "G- or J-sounding name."

When the medium gets a hit, she builds on that information. If she doesn't get a hit, she quickly moves on, making other vague references, until the sitter responds to something. The sitter will usually be eager to help, and it is easy to get hits. Nearly anyone can react to the mention of a common object such as a hat or a key, and the sitter's recollection can then be interpreted as a hit. If things aren't going well, the medium may suggest that the sitter is having trouble remembering.

Sometimes, with a show of confidence, she may suggest that the sitter write down the question and think about it later. In any case, when the medium's performance is later recalled, only the hits will be remembered.

A good psychic doing a cold reading can make surprisingly specific suggestions and get positive responses. To show how easy this is, one investigator carried out a survey in which

[m]ore than 6,000 people were asked to state whether quite specific statements were true of them. More than one third of people endorsed the statement "I have a scar on my left knee" and more than a quarter answered yes to the statement "Someone in my family is called Jack."

The results produced by "hot reading," in which the medium has advance information, can be even more impressive. Advance information is not hard to get. One technique used in stage shows is for the psychic's confederates to circulate in the audience before showtime, talking to members of the audience and eavesdropping on conversations. Then, during the show, when the psychic tells a member of the audience that her niece named Jocelyn is in the hospital for an operation, onlookers will conclude that the psychic's powers must be real. But sometimes hot reading can backfire. John Edward, a "psychic medium" who had his own TV show, appeared on NBC's *Dateline* to demonstrate his abilities. Referring to the spirits with whom he was communicating, Edward said, "They're telling me to acknowledge Anthony." When the cameraman indicated that that was his name, Edward acted surprised; but later he admitted he'd already met the cameraman and knew his name. Even then, he insisted he was merely repeating what the spirits had told him.

4.3. Hume's Argument against Miracles

When it comes to the question of life after death, many mainstream scientists may seem closed-minded and uninterested in examining the evidence. Why should that be? Part of the answer is that they have better things to do. Life is short, and once it becomes clear that an area of research is full of error and fraud, scientists naturally lose interest in it. But there is another

reason for the lack of interest, connected with Hume's argument against miracles.

David Hume, the great Scottish philosopher of the eighteenth century, did his most important work in his early 20s. In later life he gained fame for a long series of books and essays on philosophy, history, psychology, religion, and morals. Hume worked as a tutor, as secretary to the British ambassador in Paris, and as a librarian at the University of Edinburgh. But the university professorship that he desired was denied him, partly because of his scandalous views about religion and morals. One example of these views is the essay on miracles that he included in his book *An Enquiry Concerning Human Understanding* (1748).

Hume did not say that miracles can't occur, but he did argue that it is never reasonable to believe in them. Suppose, for example, someone tells you that a dead man came back to life. Should you believe it? Hume says:

When anyone tells me that he saw a dead man restored to life, I immediately consider with myself whether it be more probable that this person should either deceive or be deceived, or that the fact which he relates should really have happened. . . . If the falsehood of his testimony would be more miraculous than the event which he relates, then, and only then, can he pretend to command my belief or opinion.

In other words, we have to ask which is more likely:

- (a) That a dead man came back to life
- (b) That the report is mistaken

Of course it is more likely that the report is mistaken. "We frequently hesitate concerning the reports of others," Hume notes, because we know that often people get things wrong, make mistakes, exaggerate, and even lie. These are common occurrences. On the other hand, we have never known someone truly dead to return to life. Therefore, a reasonable person will conclude that the report is probably mistaken. This will be true any time someone claims to report a resurrection.

Hume said he hoped that this argument would be "an everlasting check to all kinds of superstitious delusion," and it

applies to the matters we have been discussing. Suppose someone says he knows of a really strong case of reincarnation. We must ask which would be more likely:

- (a) That when we die our spirits go to an astral world, where they reside for an indefinite period, until they reenter the world and invade the uterus of a pregnant woman, there uniting with the fetus she is carrying, and in the process losing the memories and personality that make us the distinctive people we are
- (b) That there is some defect in the evidence

The first possibility is incredible, while the second would be unsurprising, given our experience. Thus, a reasonable person will think (b) more likely; in fact, he or she will think (b) is *much* more likely than (a). This is the underlying reason why mainstream scientists don't want to pore over an endless amount of "evidence" concerning postmortem existence.

The Problem of Personal Identity

To capture my own existence it seems enough to use the word “I,” whose meaning is entirely revealed on any occasion of its use.

“I know what I mean by ‘I.’ I mean *this!*”

—THOMAS NAGEL, *THE VIEW FROM NOWHERE* (1986)

5.1. The Problem

Can we possibly survive death? The answer depends on what “we” are. Is a person the sort of thing that can survive the disintegration of the body? Most people in our culture seem to think so; they believe that after you die, your body rots in the ground, but you wake up in heaven (or somewhere else you don’t want to be). Does that belief make sense? Could the person who dies on earth be the *very same person* who wakes up in heaven? To answer these questions, we must grapple with the problem of personal identity. We must think about what persons are.

A Thought Experiment. Suppose scientists invent a machine that, they hope, can make a perfect copy of anything. It works by scanning an object’s atomic structure and then assembling identical atoms in the same pattern. The Duplicator successfully copies various items, starting with paper clips and ashtrays, and then a computer that runs perfectly. It makes a chocolate cake that tastes fine. When a painting by Picasso is duplicated, the scientists don’t pay close attention, and now no one knows which painting is the original. The ultimate challenge, however, is to reproduce life. Preliminary tests go well—a mouse is duplicated, and the copy is every bit as frisky as the original.

Now the scientists want to conduct one final test. They want to duplicate a human being, and they have picked you for the honor. If you agree, a copy of you will be made that will have your memories, your beliefs, your desires, and your personality. It will believe it is you, and no one—not even your closest friends and family—will be able to tell the difference. It will, for all practical purposes, *be* you. There is only one catch: We can't have two of you running around, so after the procedure is completed, the original "you" will be destroyed and the new "you" will carry on as before. For your trouble, you will be paid a million dollars.

Would you agree? One way to think about the situation is this: there is *you*, and after the test there is *you plus \$1 million*. On this view, you should take the money. Few of us, however, would think that way. Instead, we would think: *I'll be dead*. It doesn't matter that the duplicate will act like me and have my memories and emotions. Nor does it matter that the duplicate will believe it is me—the fact is, it won't be. I'll be dead, and it will merely be a copy of me.

Now consider the afterlife. According to traditional belief, when you die you will lose consciousness and your body will decay. Here on earth, it will seem like you no longer exist. Meanwhile, in another world, someone will appear with your beliefs and memories. That person will believe she is you. But will she really be you? Why won't she just be your duplicate? This question is important. All of us want to go to heaven, but few of us care whether our duplicate gets to go.

5.2. Personhood at a Time

The problem of personal identity is the problem of what we are. It has two aspects. First, there is the question of what a person consists in at a particular time. What does it mean to be you, right now? Second is the question of what makes someone the same person at different times. The child at your seventh birthday party, blowing out the candles, was the person who is now reading this sentence. How should we explain that fact? These questions—of personal identity at a time and over time—are related. We'll take them up in turn.

The Bundle Theory and the Soul Theory. What does a human being like you consist in, at a given moment or over a brief

stretch of time? Two points should be uncontroversial. First, you have a physical body. Second, you have a mental life that takes place in, and is sustained by, your physical body.

Your mental life includes perceptions, thoughts, memories, emotions, and the like, which are connected with one another in various ways. The individual items come and go. Some, such as your awareness of this sentence, are short-lived. Others, such as your memories of your parents, may persist for a long time. Some of them may be related to one another by cause-and-effect: for example, remembering the day you passed your driver's test might make you feel proud. When you go to sleep, many (if not all) of your mental items cease to exist; and when you wake up, they are succeeded by a similar parade of items, many of which resemble and are caused by the earlier ones.

Does anything more need to be added? One hypothesis is that you consist simply of your body together with this collection of mental items. This is the Bundle Theory. Many people, however, think that the Bundle Theory leaves out something important. They believe something like this:

Within each person is a kernel that may be called the *soul* or the *ego* or the *self*. It is the subject of all the person's experiences. It is simple and indivisible. And it is present throughout the person's life.

This is the Soul Theory. The Bundle Theory denies that there is any such soul. On the Bundle Theory, *the parade of mental items is all that exists*, so far as the "mind" is concerned.

The Bundle Theory may be hard to accept, because you may feel that, in addition to the various experiences you have, there is the being who *has* the experiences—and that being is *you*. David Hume (1711–1776) understood this problem. He said, "[Human beings] are nothing but a bundle or collection of different perceptions"; and when asked whether there mustn't be a "self" that has those perceptions, he replied that there is no reason to think such an entity exists:

[W]hen I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception.

Consider this analogy: A car consists of an engine, a body, a transmission, wheels, an electrical system, and so on. Suppose we listed all the parts of a car and asked: “But isn’t the car itself something separate from all these parts? After all, it is the car that *has* an engine, a body, a transmission, and so on.” Obviously, this is silly—the car is not some mysterious thing in addition to all of its parts. The car is simply the collection of those parts arranged in a certain way. Similarly, a “self” may simply be a bundle of mental items, combined with a body, and nothing more.

Split Brains. How can we decide between the Bundle Theory and the Soul Theory? Hume’s point was that the Bundle Theory is faithful to what we actually observe. We observe the parade of mental items, but we never observe a “self.” This eighteenth-century observation accords well with our present-day scientific view of the person. In a scientific account of human beings, we are apt to talk about behavior, the brain, various mental phenomena, and the relations among them; but scientists do not talk about a separate, simple “self” or “ego.” And we may add to the scientific picture an amazing discovery about the brain.

In 1981, Roger W. Sperry was awarded the Nobel Prize in Physiology or Medicine for his work on split brains. Earlier in his career, Sperry had demonstrated that different parts of the brain serve different functions. Then, in the 1960s, he and his co-workers studied what happens when the brain’s two hemispheres are split apart.

In a normal human being, the left hemisphere of the brain contains the “language center” and does mathematics and other kinds of analytical thinking. The right hemisphere specializes in other matters, such as spatial comprehension, map reading, and recognition of faces. The right side can do simple calculations—it can perform addition, but only up to about 20. The right side of the brain controls the left side of the body, and vice versa.

You might think that the two hemispheres would sometimes interfere with each other, with signals from one side contradicting signals from the other side. Normally, however, this does not happen, because the hemispheres communicate with each other through a bundle of nerves called the corpus callosum. When the corpus callosum is severed, as is sometimes

done to treat epilepsy, the hemispheres cannot communicate, and strange things happen:

- If a split-brain patient is asked to reach into a bag and identify an object, such as a pencil, she will be able to identify the pencil with her right hand but not with her left hand.
- When the words “Give me the pencil” are shown to the left side of her visual field, she will insist that she cannot see any pencil. At the same time, her left hand will be picking out a pencil from a collection of objects and offering it to you.
- Similarly, if an odor from a clove of garlic is fed into her right nostril, the patient will deny that she can smell anything; but if she is asked to point to what she has smelled, the left hand will point to the garlic.

What should we make of this? It looks as if the patient has split into two persons who are at odds with each other. Sperry himself concludes that each side of the brain is “a conscious system in its own right, perceiving, thinking, remembering, reasoning, willing, and emoting, all at a characteristically human level, and . . . both the left and the right hemisphere may be conscious simultaneously in different, even in mutually conflicting, mental experiences that run along in parallel.”

The Oxford philosopher Derek Parfit saw a connection between these results and the Bundle Theory. Parfit considered this case: A split-brain patient is shown a blue patch in the left side of his visual field and a red patch in the right side. Pencils are placed in each of his hands, and he is asked to write down how many colors he sees. The left hand writes “only one” and the right hand writes “only one.” So far, so good. But when asked what color he sees, one hand writes “blue” while the other hand writes “red.”

We have here two streams of consciousness. In one, blue is experienced and reported; in the other, red is experienced and reported. The two streams do not interact. How many persons are there under these circumstances? If we assume that a person is a soul or a “self” that underlies experience, then the question is surprisingly hard to answer. “One person” sounds wrong, because each stream seems to have its own “self”—the self associated with one stream has no access to the experience or behavior

of the self associated with the other stream. But if we say that there are two selves, we invite a barrage of embarrassing questions: Where did the extra self come from? Is one of the selves the same as the original person? If so, which one? And if neither is the original, where did the original person go? Did he die?

The split-brain case exposes the weakness of the view that people consist of souls. The Bundle Theory, however, provides a neat description of what is happening in these cases. There are now two streams of consciousness, because the two hemispheres cannot interact. Each stream contains familiar sorts of elements. That is all there is to it. Parfit concludes that the numbers of selves present in a split-brain patient is the same as the number of selves normally present in us—none.

This is unsettling. If the Bundle Theory is correct, then we are not what we thought, and we may not like what we are. Common sense says that, when we die, it is at least possible that we will go to heaven. But if the Bundle Theory is true—if we are just a body and a parade of mental items—then how could we survive the destruction of our body, the disabling of our brain, and the disruption of our conscious experiences? A belief in heaven seems to presuppose that a person is a kernel, or a soul, that could shed both body and mind.

5.3. Personhood over Time

If the Soul Theory were true, then it would be easy to say what makes you the same person at different times—the same kernel is present at each time. If that little kid at the birthday party was you, then this is because you and he have the same soul (or better yet, you and he *are* the same soul). Admittedly, it would be hard to know whether the same kernel was present at different times; the soul can't be seen or touched or detected in any normal way. However, we would at least know what personal identity over time consists in.

On the other hand, if the Bundle Theory is true—if a person is a body plus a collection of mental items—then a fundamental problem remains: Which part of the bundle is crucial for your continued existence? Obviously, you could lose your left leg and still be you, but could you survive the loss of your memories, or your brain, or your personality? Different theories give different answers.

Numerical and Qualitative Identity. There are two ways in which someone might be “the same person” at different times. On the one hand, you are the same person you used to be if you have the same personality—your character hasn’t changed, your sense of humor is the same, and so on. Suppose that my old college roommate, John Smith, was likable and quick-witted. But over the years he has become mean and humorless. I might complain that this new John Smith is not the person I roomed with; he has changed. If I said this, everyone would know what I meant. John Smith is *qualitatively* not the same as he used to be.

On the other hand, however, he *is* the person I roomed with—he is still John Smith, the guy I roomed with in college. He has not been replaced by an imposter. If the sheriff comes looking for the man who roomed with me, there he is. John Smith might have changed over the years, but that doesn’t matter—the *same person* who was likable and funny is now grim and unpleasant. He is *numerically* the same individual. When philosophers ask whether person A is the “same person” as person B, they are almost always asking whether A and B are *numerically* the same.

Is personal identity important? Does it matter whether someone in the future will be *you*? Remaining the same person qualitatively is a good thing only if you are a good person; otherwise, it would be better if you changed. But we all want there to be a future person who is numerically identical with us. Otherwise, we will not exist. What does it take to be the same person, numerically, at different times?

5.4. Bodily Continuity

There are various ideas about what makes someone the same person at different times. One is the Same-Body Theory:

The Same-Body Theory: X, who exists at an earlier time, and Y, who exists at a later time, are the same person (numerically) if and only if they have the same body.

Years ago, when I last saw my roommate, someone could have (in theory) started observing him, and they could have traced the long, unbroken path through space and time that led him into my presence again today. That path connects the person then with the person now and makes them the same. Of course,

his body may have undergone some changes. He may have gained a little weight and lost a little hair. But that does not matter. Physical objects can change, within limits, and yet remain the same physical objects.

In everyday life, the Same-Body Theory works well enough. Sameness of body and sameness of person always go together, so it is natural to think they are the same thing. But this criterion seems to rule out the possibility of postmortem existence. If my roommate dies and his body is rotting in the ground, then he cannot be in heaven. Someone else might be there, with “his” memories and personality, but it cannot be him.

The Prince and the Cobbler. Is the Same-Body Theory correct? John Locke (1632–1704), the first major philosopher to discuss personal identity, argued that “sameness of body” cannot be the measure of whether two individuals are the same person.

Suppose, Locke said, we imagine that the mind of a prince is magically transferred into the body of a cobbler—a common laborer who makes shoes. The prince wakes up one morning, looks around, and is horrified to find himself in the cobbler’s humble bed. He looks in the mirror and is even more horrified to see the cobbler’s face staring back at him. Pulling himself together, he marches off to the palace, tells the guard that he is the prince, and orders him to open the door. Meanwhile, in the palace, the cobbler’s mind has been transferred into the prince’s body. Upon awakening, he is frightened to find himself in the prince’s bed, and he tries to leave before he is discovered.

Now consider the individual who has the cobbler’s body but the prince’s memories and personality. Who is this person? Is it the prince or the cobbler? If the body is what matters, it must be the cobbler. But Locke thought otherwise: “Everyone sees he would be the same person with the prince, accountable only for the prince’s actions.” Therefore, Locke concluded, we need a different theory of personhood.

The Ship of Theseus and the Human Body. Plutarch, an ancient Greek writer, told a riddle about an ordinary ship. This ship, the Ship of Theseus, was composed of wooden planks, and once in a while a plank would get damaged or become rotten and then be replaced. Eventually, *all* of the ship’s planks had been replaced. This raises a question: Is the new Ship of

Theseus the same ship—*numerically* the same—as the old Ship of Theseus? It would seem not. A ship is composed of a bunch of planks; once the planks have been removed, the ship is gone. Gradually, it seems, the Ship of Theseus was disassembled.

Now consider the human body. It gets disassembled, too. Our bodies are composed of cells, and new cells are constantly being created to replace the old. Old bone cells are replaced by new bone cells; old skin cells are replaced by new skin cells; old heart cells are replaced by new heart cells. This process is called “cell regeneration” or “cell turnover.” After about nine years, most of the cells in the human body have been replaced. So, the body of a 60-year-old is composed mostly of cells less than nine years old.

These facts create another problem for the Same-Body Theory. If our bodies get replaced, bit by bit, then no human body lasts more than nine years or so. Thus, no human being survives more than nine years, if we are our bodies. The John Smith of today cannot be the person I roomed with many years ago because these two individuals have different bodies. This implication of the Same-Body Theory is certainly odd, if not downright absurd.

Now consider some other facts about human biology. The main type of brain cell is the neuron. Although our cells are constantly being replaced, most of our neurons never get replaced. Some do, but most do not. An old man had most of his neurons when he was an infant. Thus, even if our bodies don’t exist more than nine years, our brains do. And perhaps the persistence of the brain is enough to ensure one’s continued existence. The brain, as the seat of higher functioning, does occupy a special place in the human body. This suggests a new theory:

The Same-Brain Theory: X, who exists at an earlier time, and Y, who exists at a later time, are (numerically) the same person if and only if they have the same brain.

This theory could be defended from the Ship of Theseus example. In the actual world, our neurons do not get replaced like the planks of Plutarch’s ship. We retain the same brain throughout our lives.

This theory is promising, but we won’t try to assess it. However, I will mention one problem. When a person dies, her brain has ceased functioning in important ways. Nevertheless, her brain may still exist because it takes time for one’s body

to decompose. Suppose, for example, that a healthy person drowns over a 10-minute period. At the end of those 10 minutes, she no longer exists (on earth, anyway), but her brain is still there. Those who believe in the Same-Brain Theory will be hard-pressed to explain how she has ceased to exist while her brain is still around. In its simplest form, then, this theory fails.

5.5. Memory

A different theory of personal identity focuses on mental connections—in particular, the connections of memory. I am certain that I am the person who went to the movies yesterday, sat in the third row, and talked to his wife, because I *remember* going to the movies, sitting in the third row, and talking to my wife. I am certain that I was John Smith's roommate because I remember living with him. Where the connections of memory are strong, personal identity seems certain. So we might propose this theory:

The Memory Theory: X, who lived at an earlier time, is the same person (numerically) as Y, who lives at a later time, if and only if Y can remember doing what X did, feeling what X felt, thinking what X thought, and so on.

If this theory is correct, it explains why the individual in the cobbler's body is really the prince. He is the prince because he remembers living in the palace, giving orders to the servants, and so on. This theory also accommodates the idea of an afterlife—if the person in heaven has your memories, then that person is you.

This was Locke's theory. However, it is open to a decisive objection. We forget things all the time. I cannot remember any of the thoughts that I had an hour ago. On the Memory Theory, we cannot say that I am the person who had those thoughts. This seems absurd.

Thomas Reid (1710–1796), the Scottish “philosopher of common sense,” brings up a related objection to the Memory Theory:

Suppose a brave officer to have been [beaten] when a boy at school, for robbing an orchard, to have taken a [battle flag] from the enemy in his first campaign, and also to have been made a general in advanced life; suppose, also, which must be admitted to be possible, that, when he took the [flag], he was conscious of his having been [beaten] at

school, and that, when made a general, he was conscious of taking the [flag], but had absolutely lost the consciousness of his [beating].

This is a common turn of events. When a man is 25, he may remember being 12; and when he is 60, he may remember being 25 but not remember being 12. If we apply the Memory Theory to this scenario, we get a strange result: The boy is the same person as the young man, and the young man is the same person as the general, but the boy is not the same person as the general.

What is wrong with this? The problem is that the relation “same person as” is *transitive*. This means that if A is the same person as B, and B is the same person as C, then A must be the same person as C. For example, if Lee Harvey Oswald is the person who killed John Kennedy, and the person who killed John Kennedy is the person who was shot by Jack Ruby, then it follows that Lee Harvey Oswald is the person who was shot by Jack Ruby. On the Memory Theory, however, it turns out that the same-person relation is not transitive: A (the boy) might be the same person as B (the young man), and B (the young man) might be the same person as C (the general), yet A (the boy) is not the same person as C (the general). Therefore, the Memory Theory cannot be right.

However, the Memory Theory can be revised to get around these problems. Rather than ask whether the later person *remembers* the earlier person’s life, we can ask whether a *chain of memories* connects the two individuals. When Y remembers X’s life directly, then the chain is short: Y and X are the same person because Y remembers being X. However, in other cases the chain may be longer. In Reid’s example, the young man connects the general to the boy. The general is the same person as the young man because the general remembers being the young man; the young man is the same person as the boy because the young man remembers being the boy; and so, the general is the same person as the boy. In my earlier example, though I don’t remember what I was thinking an hour ago, I do remember what I was thinking five seconds ago; and five seconds ago I remembered what I was thinking ten seconds ago; and so on. Thus, we have:

The Memory-Links Theory: X, who lived at an earlier time, is the same person (numerically) as Y, who lives at a later time, if and only if a chain of memories links X and Y.

This theory has an interesting implication about dreams. Most of us have forgotten our dreams by the time we wake up in the morning. Does this mean that we were not the ones who had the dreams? According to the Memory-Links Theory, the question is whether a chain of memories can link “Dream Person” to “Morning Person,” thus establishing their identity. When you dream, you sometimes dream about real events or real people. If so, a chain can be found: Dream Person and Morning Person may both remember something that happened before the dream, to Earlier Person. If, however, Dream Person remembered nothing about Earlier Person, then no link would exist. In that case, Morning Person would not have been the one who had the dream. Rather, Dream Person would have existed only during the dream.

Memory and Responsibility. Locke favored the Memory Theory not only because it explains how we might survive death but also because it fits with our beliefs about responsibility and punishment. In discussing the prince and the cobbler, Locke notes that the individual with the prince’s memories should be held accountable for the prince’s past actions; and from this he concludes that the individual with the prince’s memories *is* the prince, even though he has the cobbler’s body. Personhood follows the memories, not the body.

Let’s consider Locke’s argument in its most general form:

- (1) Memory always goes with responsibility. (Whether you should be held responsible for an action depends on whether you can remember doing it.)
- (2) Responsibility always goes with identity. (You are always responsible for what you do; you are never responsible for what others do.)
- (3) Therefore, memory always goes with identity.
- (4) Therefore, some version of the Memory Theory is true.

What are we to make of this? In defense of the first premise, responsibility does seem to depend on memory. Suppose my old roommate, John Smith, committed a crime—say, he robbed a bank—but then he suffered amnesia. When he is now shown the surveillance video of the robbery, he is amazed. In these circumstances, it would seem strange to blame him. Because he

cannot remember committing the crime, he no longer seems to be a proper target of our indignation.

However, this line of thought is too weak to justify the first premise. It does not show that memory *always* goes with responsibility, as the premise states. Suppose that, after many years in prison, a murderer serving a life sentence has changed—perhaps he has become religious, or has written a remorseful autobiography, or has simply matured. If we are convinced that the transformation is genuine, we may conclude that it is no longer appropriate to keep him behind bars. The prisoner *remembers* committing the act, but he is no longer responsible for having done it. In this scenario, memory doesn't go along with responsibility. Also, responsibility does not go along with identity, in violation of the second premise: The prisoner is the person who committed the crime, but he is no longer responsible for it.

The idea here is that responsibility seems to depend on sameness of person in the *qualitative* sense only. After his transformation, the murderer is a “different person” in the sense that his character has changed fundamentally. That is why it seems wrong to keep him locked up. We have “statutes of limitation” on many crimes partly for this reason: If someone shoplifted in 1979, we might wonder whether she is still the “same person” today, or whether her character has changed significantly in the intervening years. None of these thoughts need affect our belief that the reformed individual is the person who committed the crime—the *numerical-identity* claim holds true.

Is the Memory Theory Trivial? Memory is notoriously unreliable. We “remember” things that never happened, and we forget things all the time. Even when our memories are mostly accurate, our minds will add in fictitious details. For example, people who were alive at the time almost always believe that they remember when President Reagan got shot, or when Princess Diana died, or when the World Trade Center collapsed. Psychologists call these “flashbulb memories.” Are they accurate? In 1986, the space shuttle *Challenger* exploded on live TV, killing everyone on board. The next day, the psychologist Ulric Neisser asked a group of people to write down what they had been doing when they heard the news. Two and a half years later, when Neisser asked them again, they expressed great confidence in their memories, but only 3 people out of 44 gave

answers that were accurate in their details. Meanwhile, over half of the subjects gave answers that were simply wrong. Many of our “memories” are like this.

The Memory-Links Theory says that X and Y are the same person if a chain of memories connects X and Y. But what do we mean by a chain of *memories*? We must distinguish real memories from apparent memories. Real memories are accurate representations of things that happened to us, while apparent memories include the false “recollections” that sometimes mislead us. When Y remembers doing what X did, is the memory real or apparent? Obviously, the theory cannot be referring to apparent memories. If it were, then any lunatic who “remembered” writing *Hamlet* would be William Shakespeare. So the theory must be referring to real memories—X is the same person as Y if a chain of *accurate* memories links X to Y. This seems correct, but only because it is trivial. If Y accurately remembers being X, then of course X and Y are the same person—this follows from the idea of “accurately remembering.” The Memory-Links Theory thus explains “same person” in terms of “accurate memory,” and it explains “accurate memory” in terms of “same person.” Thus, it cannot help us understand what persons are. Bishop Joseph Butler noticed this problem in 1736: “[O]ne should really think it self-evident,” he wrote, “that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity.”

Conclusion. The Same-Body Theory and the Memory Theory are the most obvious attempts to define personal identity over time. However, there are others. We have already mentioned the theory that emphasizes the brain. Other theories focus on personality, on causal connections, or on some combination of features. Different philosophers favor different approaches.

What is most striking in all this is what philosophers today seem to agree on. They agree that the Bundle Theory, not the Soul Theory, is the correct view of personal identity at a time. And so they think that identity over time has to do with the continuation of certain bodily or mental features, not with the ongoing existence of a soul, or an ego, or a self, or a kernel.

If these philosophers are right, then we might have to reevaluate our attitudes about life and death. Right now we care far more about ourselves than about others, and we regard our

own deaths with horror. These attitudes seem to go along with the Soul Theory: If I am a kernel—if my identity consists in a simple and indivisible entity that is *me*—then naturally I will care especially about that kernel, and I will not want it snuffed out for all eternity. However, on the Bundle Theory, both my life and my death appear different. I am just a body combined with certain mental states. There is nothing special about what I am, except for the trivial fact that no two people are exactly alike. When I die, nothing magical will happen. In the future there will be other minds and other bodies, but none of those minds and bodies will bear the right relation to me for it to be true that I am one of them.

As Derek Parfit has noted, these changes in our outlook might be liberating. We might care more about others, and we might worry less about our own mortality. Meanwhile, the question of whether the individual in heaven is you or is only your duplicate may suddenly seem trivial. There are a limited number of relevant facts: We know that you, on earth, have various thoughts, memories, and so on; and we know that the individual in heaven has similar thoughts, memories, and so on. This is all there is to know. If we ask whether the individual in heaven is *you*, and we mean to be requesting some further information, we are going to be disappointed. The person in heaven has no soul, and neither do people on earth.

Body and Mind

How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djinn when Aladdin rubbed his lamp.

—T. H. HUXLEY, *LESSONS IN ELEMENTARY PHYSIOLOGY* (1866)

6.1. Descartes and Elizabeth

Like Socrates, René Descartes (1596–1650) believed that the body and the soul are different kinds of things. Using traditional terminology, Descartes said that they are different *substances*. The body is a material substance. It is like a machine, with parts that work together according to the laws of physics; and, like a machine, a body is incapable of thought and feeling. The soul, on the other hand, is an immaterial substance. It weighs nothing and has no measurable dimensions. But it does have thoughts and feelings. One's entire mental life is the life of the soul. A human being is therefore a compound entity, a combination of a body and a soul.

Mind–Body Dualism. The mind–body problem arises because there seem to be two radically different kinds of facts about human beings. On the one hand, there are *physical facts*: A woman may be five feet tall, with green eyes, a heart, a brain, and two big toes. On the other hand, there are *mental facts* about her: She has beliefs, desires, and intentions. The problem is to explain the nature of the mental facts and their relation to the physical facts.

Mind–Body Dualism is one attempt to solve this problem. This theory says that the mind and the body are different kinds of things: The body is physical, while the mind is nonphysical. So, physical facts are facts about the body, while mental facts are

facts about the mind. To prove that the body and mind are different, Descartes observed that we can imagine the body without the mind, and the mind without the body. We can imagine a human robot—someone who was physically like us but had no consciousness or inner life. And we can imagine spirits—creatures with emotions and desires but no bodies. If two things can exist without each other, then they are not the same thing. Thus, body and mind are distinct. This has been called The Conceivability Argument for dualism.

Descartes also emphasized two differences between mental states and physical states. First, he said, mental facts and physical facts are known in different ways. Physical facts are known by observation and experiment. Thus, they are *publicly available*. Any one can look at you and see how tall you are. You yourself have no special access to such information. You learn your height in the same way that everyone else does, by looking and measuring. In fact, other people may know more about your body than you do. A biologist may know that your heart has four chambers, even if you don't know that. Your doctor may know that your blood has a high sugar content, while you know nothing about it.

Mental states, by contrast, are *private*. Each person has special access, through introspection, to his or her own thoughts. You know that you are thinking about George Clooney simply because you are aware of having those thoughts, while other people will know only if you tell them. You have privileged access to your own mental states.

Second, Descartes said, you know your own mental states with greater certainty than you know any physical facts. We can easily be mistaken about physical facts, including facts about ourselves. A woman might believe she is five-feet-one when really she is only five feet. You might believe that you have tonsils, because you have forgotten about the operation you had as a child. Of course, we may believe some things about ourselves with a high degree of confidence. You may feel certain, for example, that you have two arms. But, at least in principle, you could be wrong about even that. You could be crazy, or, as Descartes says, an evil demon could be deceiving you.

But, once again, the story about mental facts is different. Each of us, Descartes said, is *infallible* when it comes to the contents of our own minds. If you think you are feeling pain, then you are. You cannot be mistaken, even in principle. (Imagine that,

when you tell your doctor you need a painkiller, she replies: “I’m sorry, but you’re mistaken—you aren’t really feeling anything.” That would be crazy.) Similarly, if you sincerely believe you are thinking about a handsome middle-aged man, then you are.

Of course, you may be mistaken about the causes of your mental states. You may believe you are hearing a live guitar when in fact you are only hearing someone play Guitar Hero. But *that you are hearing sounds of a certain nature* is beyond doubt. Similarly, you may have false beliefs about George Clooney—you may be thinking of the star of a recent James Bond movie, believing it was Clooney, when really it’s Daniel Craig. But *that you are thinking of a man with a certain appearance* is undeniable. Descartes’ point was that we can never be mistaken about what our current experience is like.

Elizabeth’s Objections. Born into a well-to-do French family, Descartes took a law degree when he was 20 and then joined the army of Bavaria, which, he figured, would give him free time for his intellectual work. After a couple of years he quit the army and moved to Paris. Later, when his writings became controversial, he moved to Holland, a country with a long history of providing safe haven for radical thinkers and persecuted religious minorities. There he became a favorite of royal ladies. In 1649, Sweden’s Queen Christina invited him to Stockholm to teach her philosophy. In Stockholm he died of pneumonia, brought on, legend has it, by having to give her lessons at 5 A.M.

Another royal lady in Descartes’ life was Princess Elizabeth of Bohemia (1618–1680). Elizabeth was a gifted woman who might herself have become a great philosopher, had social customs been different. As it was, she spent her life in genteel poverty, and in later life she became a nun in a Protestant convent. She never married, she said, because she preferred intellectual pursuits. In 1642, she read Descartes’ *Meditations on First Philosophy* and began a correspondence with him that lasted until his death. In these letters she identified one of the main difficulties with Mind–Body Dualism.

The problem is that it’s hard to see how a nonphysical mind could interact with a physical body. It is obvious that mental events can cause physical events—you decide to do something, and your body responds. It is equally obvious that events in your body cause events in your mind—when you step on a

tack, you feel pain. But if the body is physical and the mind is nonphysical, how is this interaction possible? In a letter to Descartes, Elizabeth noted that, for a body to be moved, it must be pushed. But a nonphysical mind cannot push anything, because it lacks extension. So how is this interaction possible?

Descartes did not have a good answer. He wrote back and suggested, somewhat lamely, that the mind might work like gravity. At that time, before Newton, gravity was thought of as something inside objects that somehow causes them to move toward the center of the earth. Similarly, Descartes thought, the mind might somehow cause the body to move.

Elizabeth was unimpressed. "I cannot see," she replied, "why this should convince us that a body may be impelled by something immaterial." Instead of thinking of gravity as an "immaterial cause," she said, we should search for a better understanding of what gravity is. She also expressed puzzlement about what "immaterial substances" are supposed to be. She complained that immaterial substances are always defined in terms of what they are *not*: An immaterial object has no mass, it does not occupy space, and so on. But what *is* it? Elizabeth was understandably perplexed.

Faced with such a perceptive critic, Descartes was out of ideas. With nothing further to offer, he concluded the exchange by assuring the princess that the subject was very difficult and telling her that she should not spend too much time worrying about it.

Radical Emergence. Another problem for Mind–Body Dualism is to explain why the immaterial mind exists at all. In the womb, a human being begins as a purely physical thing. But then one day, after the brain has reached a certain level of complexity, poof!—a new kind of substance pops into existence. Why? And remember that human beings are not the only conscious animals. Life existed on earth for hundreds of millions of years before a species evolved that could think and feel. So, for millions of years, there were only physical systems on earth. And then one day, seemingly out of nowhere, consciousness comes on the scene. Why? This is called *the problem of radical emergence*. On a Dualist perspective, as Huxley said, the emergence of mind from matter seems inexplicable, like the emergence of the genie from the lamp.

Despite these objections, Mind–Body Dualism remained popular well into the twentieth century. The theory finally lost favor because it did not fit the emerging scientific picture of human beings. As our understanding of human beings progressed, Descartes’ immaterial substances came to seem more and more like prescientific, superstitious entities, like ghosts. As a result, Mind–Body Dualism has largely disappeared from science and philosophy.

6.2. Materialist Theories of the Mind

A human being is a vast collection of atoms, organized in a complex way, obeying the laws of physics, chemistry, and biology. Materialism claims that this is *all* we are—we have no ghostly, nonphysical parts. Materialism is, of course, a controversial doctrine, and the debate over Materialism is often seen as a debate about religion. Here, however, the issue is not religion but the nature of mental facts. Materialist theories attempt to explain those facts in purely physical terms. Is any such theory defensible?

There are two ways to proceed. A materialist theory could focus on what we can observe about people from the outside—their speech and behavior. Or it could focus on what goes on inside the person—occurrences in the person’s brain. Each of these possibilities has been tried.

The First Materialist Theory: Behaviorism. Can we explain the nature of thoughts and feelings purely in terms of behavior? Intuitively, this seems to get things backwards—instead, it seems right to explain behavior in terms of thoughts and feelings. When someone shouts at us, we say it is because she is angry; when someone grabs a sandwich, we say it is because he is hungry.

For ordinary purposes, such explanations may do just fine. However, according to a school of thought known as Behaviorism, which flourished during the middle of the twentieth century, such “mentalist” explanations have no place in a scientific account of human beings. Psychologists John B. Watson (1878–1958) and B. F. Skinner (1904–1990) were the leaders of this movement.

Science, they said, deals with what is publicly observable, and private mental states do not qualify. So, rather than saying “Jane shouted because she was angry,” behaviorists would focus

on Jane's situation and the events that triggered her outcry. For example, she shouted because she was surrounded by noisy children, and one of them squirted water in her ear. According to Watson and Skinner, the science of psychology looks for patterns of stimulus and response—physical responses (shouting) to physical stimuli (noise, water in the ear)—and formulates general laws about how behavior is shaped by physical inputs. Private mental episodes are irrelevant.

Watson and Skinner did not deny that private mental states exist; they said only that such states have nothing to do with science. But it was only a small additional step to say that private mental states do not exist, and philosophical behaviorists such as Gilbert Ryle (1900–1976) took that step. Ryle's book *The Concept of Mind* (1949) was, for a brief time, the most widely discussed book of philosophy in the English language. Ryle argued—or anyway, he has been interpreted as arguing—that so-called mental states are really nothing more than behavior. Some mental states consist in patterns of current behavior: To be angry is to be aggressive, to shout, to use harsh words, to strike out, or to turn one's back on someone. (There is a wide range of behavior associated with anger, rather than one specific behavior. That is typical.) Other mental states consist in a tendency to behave in certain ways on suitable occasions. You may want a donut, for example, even though you are currently doing nothing about it. Nonetheless, if you want it, you must be disposed to take it when it is offered, and (if you want it badly enough) you must be disposed to celebrate like Homer Simpson when you get it. Similarly, to be afraid of something is to be disposed to tremble and flee from it when you are in its presence. So Jane's being angry does not cause her to shout; rather, her anger *consists in* her shouting. This version of Behaviorism is called *Analytical Behaviorism* because it tries to analyze "mental states" in terms of behavior.

Ryle ridiculed Descartes' view as "the dogma of the Ghost in the Machine," and the image stuck. By exorcising the ghost, behaviorists avoided the problems that had plagued Mind–Body Dualism. Gone were the problems of explaining what nonmaterial substances are, how they come to exist, and how they interact with physical bodies. These were problems for those who believe in ghosts.

The ideas of privileged access and infallibility also fell by the wayside. The behaviorists noted that an outside observer

can often see that someone is angry, or afraid, or in love, just as easily as, or perhaps even more easily than, the person himself. As for infallibility, only a little reflection is needed to realize how often we are wrong about our own mental states. Someone who is depressed, or afraid, or resentful, will often think otherwise; and what you want may sometimes be obvious to other people, even if you don't know it. So it looks as if our "mental" states are in fact publicly observable, which makes it unnecessary to understand them as private inner episodes.

The Failure of Behaviorism. Behaviorism was appealing because of its no-nonsense, scientific stance. Nonetheless, after enjoying almost a half-century of popularity, it was abandoned by psychologists and philosophers. In psychology, the behaviorist research program was replaced by a variety of investigative methods that went under the heading of "cognitivism," while in philosophy, Analytical Behaviorism could not be defended against a number of objections.

First, Analytical Behaviorism seems false on its face. At any given moment, each of us is aware of having perceptions, thoughts, and so on, even if no behavior is occurring. While sitting in my armchair, all sorts of mental activity may be going on: I can see the green wall in front of me; I can feel the chair's fabric beneath my hand; and all the while I am thinking about how no one except a few astronauts has ever seen the dark side of the moon. It just seems false to say that this mental activity consists of nothing but behavior or possible behavior. The stream of consciousness seems to have a reality of its own, independent of its connection (or lack of connection) to behavior.

Second, not every mental state can be understood as a pattern of behavior or as a disposition to behave. So long as we concentrate on examples such as fear and desire, the behaviorist analysis seems plausible. But consider a different sort of example, say, the experience of listening to Chopin's Nocturne in C Minor—what pattern of behavior is this sublime experience supposed to be? And how is this behavior different from the behavior associated with listening to Bach's Mass in B Minor? To put the matter even more simply, think of how a major chord sounds and how a minor chord sounds. Is the difference between them supposed to be a difference in behavior? When you consider these kinds of examples, Behaviorism seems far-fetched.

Third, it seems possible for two people to behave in exactly the same way and yet have different mental states. In the seventeenth century, John Locke suggested that one person's sense organs might differ from another's, so that in the same circumstances, one person sees red and the other sees blue. In stopping at a traffic light on a clear day, one person would see the sky as blue and the light as red, while the other would see the sky as red and the light as blue. However, their behavior would be indistinguishable—they would both *say* “That’s red” when looking at the light, because each of them would have learned to call that color “red.” This example is known as the “inverted spectrum.” If this happened, their visual experiences would be different, but their behavior would be the same. This seems to show that, whatever mental states are, they are not behavior.

With the failure of Behaviorism, scientifically minded philosophers looked for an alternative account of mental facts.

The Second Materialist Theory: Mind–Brain Identity. Wilder Penfield (1891–1976) was a Canadian doctor who performed surgery under extraordinary conditions: he would operate on people's brains while they were wide awake. As Penfield knew, this can be done painlessly, because the brain itself has no pain receptors. Penfield began performing these operations in the 1930s in order to treat patients with severe epilepsy. Epileptic seizures, he thought, must be triggered by events in a diseased part of the brain. Therefore, if he could remove the diseased bit, perhaps he could stop the seizures. But how would he know what to cut out? His patients reported seeing lights and hearing sounds at the onset of the seizures. So Penfield tried to find the part of the brain that caused those experiences.

Thus, while his patients were fully conscious, Penfield would use an electric probe to stimulate their brains, and the patients would report what they were experiencing. One patient said he was sitting in a railway station. Another said he was snatching a stick from the mouth of a dog. Others said they were watching a circus wagon or hearing the music from the Broadway show *Guys and Dolls*. These patients did not say they were *remembering* these experiences; they said they were *having* them. Penfield's probes could also produce such feelings as fear and loneliness. In each case, the experience could be repeated by restimulating the same spot in the brain.

Penfield was not trying to prove any philosophical theory, but his results strengthened the conviction that, in some important sense, the mind *is* the brain. Penfield seemed to have identified where particular conscious experiences are located. Today we sometimes see the identity of the mind and the brain asserted as one of the main findings of twentieth-century neuroscience. But should this identification be taken literally? Many philosophers think so. Mental events, they say, are not ghostly; nor are they behavioral. Instead, they are neurological. Each of a person's mental states is identical with the firing of particular neurons, or cells, in the brain. This is called the *Mind–Brain Identity Theory*.

To see how this theory works, consider an example such as *being in pain*. Pain is a good example of a mental state: It feels a certain way, subjectively, and, as Descartes said, each of us is aware of our own pain in a way that others are not. But what is pain? Note, first, the connection between pain and behavior. If someone is in pain, she will be disposed to flinch, tighten her muscles, wince, cry out, and take action to stop the pain. She will also be disposed to say things like “That hurts” or to respond affirmatively when asked if it hurts. “Being in pain” may thus be defined as *the state of the person that causes that sort of behavior*. But what, exactly, is that “state of the person”? This is where science comes in. Science tells us that pain-related behavior is caused by particular neurological events. Therefore, we may conclude, pain *is* those neurological events. And we may hypothesize that other mental states are identical with other neurological events, even if we don't know which they are.

Of course, mental states do not *seem* like brain states. Being in pain, listening to Mozart, or thinking about Dallas seem very different from neurons firing. But it is common in science to discover that a familiar kind of thing actually consists in something unfamiliar. Water, for example, is H₂O; lightning is an electrical discharge; and heat is molecular motion. Those who believe in the Mind–Brain Identity Theory thus have a ready answer to Descartes' Conceivability Argument. The idea that the mind and brain could exist without each other, they say, is an illusion. It is like the belief that water could exist without H₂O, or that lightning could occur without electrical discharge, or that heat could exist without molecular motion. It couldn't happen.

Development of the Theory. We need to distinguish *tokens* from *types*. Suppose you and I each have a \$20 bill in our pockets. We both have the same type of thing—a \$20 bill—but we have different tokens: the particular bill (the “token”) in my pocket is not the particular bill (or “token”) in your pocket. Consider some further examples: *The Godfather* and *Casablanca* are tokens of the type “movie”; the Chrysler Building and the Empire State Building are tokens of the type “skyscraper”; Everest and Kili-manjaro are tokens of the type “mountain.”

“Neuron firing” is also a type that has different tokens. The neuron firing that occurs in my brain at a particular time is one token of this type, while the neuron firing that occurs in your brain at a different time is another token of this type. The same goes for pain: The particular pain that I am feeling now is one token; the pain that you are feeling is another. Now, with this distinction in mind, let us look again at the Mind–Brain Identity Theory. It turns out that there are two different possible theses:

1. *Type–type identity*: This is the theory that being in pain (considered as a type of thing) is identical with a particular sort of neuron firing (considered as a type of thing). It would follow from this that no one could be in pain—no token of pain could exist—unless neurons were firing in some particular way.
2. *Token–token identity*: This is the theory that each instance of pain is identical with some particular physical state. For example, the pain that I am feeling now is identical with the neuron firing that is occurring in my brain now, while the pain you felt yesterday was identical with the neuron firing that occurred in your brain yesterday.

Type–type identity is implausible for two reasons. First, pain comes in many forms: there are backaches, headaches, pinches, cramping pains, sharp pains, stinging pains, burning pains, and so on. It is unlikely that just one sort of neuron firing is associated with all of them. Second, suppose there was a life-form on another planet with a physiology very different from ours. According to the type–type identity theory, they could not feel pain. But that seems wrong. Why couldn’t pain have some other physical basis? Perhaps, for the alien beings, pain would

consist in the activity of particles that are very different from our neurons. Type–type identity theory thus seems committed to an overly narrow view of the mind—a view according to which only human beings (and other similar animals) can have mental states.

Token–token identity, however, does not have these problems. First, it is not committed to the idea that pain must always be identified with one type of thing. Rather, different tokens of pain may be identified with different kinds of things. And so, the theory does not rule out the possibility of pain in creatures with physiologies very different from ours. The pain that an extraterrestrial being felt on a certain occasion could be identical with an instance of alien brain-particle activity that took place on that occasion. Thus, most philosophers understand the Mind–Brain Identity Theory as a theory about the identity of *tokens*: the pain that I am now feeling is identical with the brain activity that is now occurring in my head.

Advantages of the Mind–Brain Identity Theory. The Mind–Brain Identity Theory was a clear advance over both Mind–Body Dualism and Behaviorism. First, the Mind–Brain Identity Theory explains the nature of mental facts without any reference to ghostly “nonmaterial” entities. Mental facts are facts about the brain. Second, there is no problem in explaining how mental events can cause physical events, and vice versa. Because mental events are brain events, all interactions between mind and body are interactions between brain and body. Mental causation is thus merely one type of physical causation. Third, the theory can explain how thought occurs without any behavior occurring. Things can be happening in your brain—neurons fire as you think of Dallas—even though there is no outward behavior.

Finally, the Mind–Brain Identity Theory helps us understand how introspection works. If mental events are occurrences in the brain, then introspection—which, after all, is just another mental event—is also an occurrence in the brain. Introspection is an “inner” process because its neural firings are not connected with inputs from the eyes and ears and other sense organs. Instead, they are connected internally with other parts of the brain. As a result, introspection gives us a special kind of access to our own mental states.

For all these reasons, the Mind–Brain Identity Theory seemed right. Here was a plausible, scientifically respectable view, free of the problems that plagued other theories.

A Third Theory: Functionalism. But suppose we now ask: Why do we call this neuron firing “a pain” rather than “a belief” or “a thought”? To answer this question, we would need to explain why the neuron firing counts as being the sort of thing that pain is. And so, we would need to explain what pain is, *considered as a type of thing*. But the Mind–Brain Identity Theory cannot do this. By focusing only on tokens, it provides no understanding of what pain, in general, is.

These thoughts pointed the way to a new theory of the mind, called *Functionalism*. Functionalism identifies pain (considered as a type of thing) with the activation of certain physical features of systems. However, it does not care what those physical features are composed of. Pain is therefore identified with *whatever* physical features of a system serve the appropriate functions within that system. So, according to Functionalism, *pain is any feature of a physical system that serves the same functions, in that system, as the neuron firings serve in us when we are in pain*. (Remember that pain is just an example; similar accounts are also given for other mental states.)

What, exactly, are the functions that the neuron firings serve in us when we are in pain? We may think of the functions in terms of environmental inputs, behavioral outputs, and relations to other mental states. The inputs are the physical events that impinge on your sense organs, causing you pain—for example, a hammer falls on your foot. The outputs include your speech, behavior, and other physical reactions to pain—for example, you wince, grab at your foot, and curse. These are the behavioral manifestations of pain. The relations to other mental states include the fact that pain makes you feel irritated, and pain makes it hard for you to concentrate on other things.

We can now say more about what pain is. According to Functionalism, pain is identical with any feature of a system that serves these functions—that links together these kinds of inputs, outputs, and other internal states. For you, therefore, a particular token of pain may be a neural firing, while for an extraterrestrial a particular token of pain may be the activation of a different internal mechanism. The physical character

of the mechanism doesn't matter. All that matters is that the mechanism serves the appropriate functions. I should add that, strictly speaking, the mechanism does not *have to be* physical. A nonphysicalist Functionalism is possible: Mental events could be nonphysical, provided they play the appropriate roles within the overall system. For example, nonphysical pain could be characterized by its relation to environmental inputs, behavioral outputs, and relations to other nonphysical mental states. But I shall continue to assume that the most sensible version of Functionalism is a form of Materialism.

Functionalism has joined the Mind–Brain Identity Theory as a leading theory of the mind. Many philosophers accept both theories: they think that tokens of human pain are identical with events in human brains; and pain, in general, is understood in terms of its functional roles.

A Problem with Functionalism. Functionalism implies that any sufficiently complex system will have a mind and that the physical composition of the system doesn't matter. The human system is made of flesh and blood, but the flesh and blood are not essential—we could be made of other stuff, like the extraterrestrials, and still be “thinking things.” We could even be made of silicon chips, like a computer.

For this reason, researchers in the field of artificial intelligence (AI) are strongly attracted to Functionalism. Functionalism supports the idea that the mind may be understood as a type of digital program. The relation between body and mind, it is said, is like the relation between hardware and software. A computer is an inert, lifeless thing until a program starts running. Then it accepts inputs, performs operations on them, and produces outputs. A human brain works in the same way: It accepts inputs, performs operations, and produces outputs. Thus, what we call the mind is simply the program running in the brain. Of course, human minds are enormously more complicated than any computer program yet written. Nonetheless, if a self-correcting, “learning” program running on a computer operated at the same level of complexity as our brains, it would have the same mental states that we have.

Although many people like this implication of the theory, it is also a source of weakness. Functionalism says that the physical composition of a system doesn't matter; all that matters

is how it operates. Is that true? Suppose we imagine a system consisting of a vast collection of beer cans connected by wires and pulleys and powered by windmills, so that the beer cans will clank against one another in a way that parallels the firings of neurons in someone's brain. Would this contraption be conscious? Functionalism implies that it must be, but critics of Functionalism think this is absurd.

6.3. Doubts about Materialist Theories

There are two problems that threaten all materialist approaches to the mind.

Subjectivity. Think of “what it is like” to see a green lawn or to taste a banana. Don't think of anything that merely accompanies the experience, like playing golf or being at the breakfast table, but think of the experience itself: the way green looks and the way bananas taste. These experiences have a subjective character; there is something that it is like to see that color and taste that fruit. More generally, there is something that it is like to be you, to move through the world perceiving it from your point of view, experiencing the sights, sounds, smells, touches, and emotions that make up your life. This is what makes you so different from a mere physical object, such as a brick. There is nothing it is like to be a brick. It would be impossible to see the world from the brick's point of view, because the brick has no point of view.

Having experiences is the essence of having a mind. Therefore, above all else, we should expect a theory of the mind to tell us about having experiences. Yet the materialist theories we've been considering can't seem to explain why there is something it is like to have experiences. The theories talk about behavior, brain states, and the functioning of complex systems. How can a brain state or a computational process be the flavor you experience when you taste a banana? Saying they're identical sounds more like a statement of the problem than a solution to it.

The philosopher Frank Jackson (1943–) put the point in this way. Suppose that a brilliant scientist named Mary has lived her entire life in a black-and-white environment. Up to now, Mary has seen only whites, blacks, and greys. Also suppose that Mary knows everything there is to know, in physical terms, about color perception. She knows, for example, what happens

in your brain when you see the color red. Then one day, for the first time, Mary sees a ripe tomato. When this happens, she learns something: she learns *what it is like to experience redness*. This knowledge is new to her, even though she already knew all the physical facts about color. Thus, “what it is like to experience redness” cannot be physical. Jackson concluded from this that Materialism is false.

Could some new materialist theory explain the nature of subjective experience? Or could some existing materialist theory be modified or expanded to explain subjectivity? Perhaps this could happen, but the problem cuts very deep. None of the available physical facts—facts about behavior, neuron firings, and the rest—seems to be the *kind* of thing that could be the look of green or the taste of a banana, because the latter have a subjective character that the former lack.

Intentionality. Materialist theories also seem unable to accommodate what Franz Brentano (1838–1917) called the “intentionality” of mental states. Brentano was a German philosopher whose best-known work is *Psychology from an Empirical Standpoint* (1874), in which he addresses the nature of the mind.

Mental states, Brentano says, are always *about* things other than themselves. My desire to see Dallas is about Dallas; your belief that seawater is salty is about seawater; the math student’s bafflement is about the square root of minus 1. This simple feature of mental states poses a formidable problem for materialist theories of the mind. How can a physical state, considered simply as a physical state, be “about” anything? How can a group of firing neurons be about Dallas, seawater, or the square root of minus 1?

There are a couple of ways in which we might try to answer this question, but neither works very well. One idea is that a physical object can be about something by *resembling* it. A picture of Taylor Swift, for example, is “about” Taylor Swift because it looks like her. But this idea does not get us far, for several reasons.

First, resemblance is a symmetrical relationship—the picture resembles Swift, but Swift also resembles the picture. Yet Taylor Swift is not “about” the picture. So the relation of “being about” cannot be the same as the relation of “resembling.” Second, brain states do not look like Dallas, or seawater, or Taylor Swift, or much of anything else. To the naked eye, the brain looks like squishy reddish grey stuff; under a microscope, a

neuron looks like a little clump with lots of stringy appendages. If “aboutness” consisted in resemblance, then neurons could be about lightning bolts or bushes, but not much more. Third, we can think about the square root of minus 1 and have beliefs about it, but no physical state could resemble it, because the square root of minus 1 does not look like anything. The notion of resemblance, therefore, is no help.

Another idea, popular among philosophers, is that brain states can be about objects if the brain state is *causally related* to the object in some appropriate way. For example, a certain state of my brain might have been caused by my seeing Taylor Swift in concert, and that is what makes it about her. Or, my brain state might have been caused by my seeing an image of Taylor Swift on TV—and the TV image was ultimately caused by a camera that was trained on Taylor Swift herself—and so, again, one could say that my brain state was about Taylor Swift.

The evaluation of this idea involves some complex issues; nonetheless, there is one reason why all causal theories of “aboutness” are implausible. Many of our thoughts are about things that do not exist—vampires, Martians, Sherlock Holmes—and nonexistent things cannot participate in causal chains. Moreover, other objects of thought, such as the square root of minus 1, could not possibly enter into causal relations, even if they are in some sense real. So, causation cannot be the whole story of “aboutness.”

It does not look as if physical states could be intentional. This seems to show that, whatever mental states are, they are not states of the brain. But we should not be too eager to draw this conclusion. Intentionality is a puzzling notion. It is the subject of ongoing debate and study, and no one has produced an account of it that satisfies the majority of investigators. Moreover, intentionality is not just a problem for Materialism; it is a problem for any theory of the mind. If you believe, like Socrates and Descartes, that the mind is immaterial, you still face the problem of explaining how fluctuations in an immaterial substance can be about Taylor Swift, or about Dallas, or about a vampire. The problem would be just as difficult for you as for the Mind–Brain Identity theorists. A more prudent course, therefore, might be to await further developments. For now, no one knows how to solve the mind–body problem.

Could a Machine Think?

Philosophers have been dreaming about AI for centuries.

—DANIEL C. DENNETT, *BRAINCHILDREN* (1998)

7.1. Brains and Computers

In 1637, René Descartes thought about whether machines could think, and he decided they could not. The mark of intelligence, he said, is the ability to understand language, and machines could never do that:

It is indeed conceivable that a machine could be made so that it would utter words, and even words appropriate to the presence of physical acts or objects which cause some change in its organs; as, for example, if it was touched in some spot that it would ask what you wanted to say to it; if in another, that it would cry that it was hurt, and so on for similar things. But it could never modify its phrases to reply to the sense of whatever was said in its presence, as even the most stupid men can do.

Descartes was right about the machines of his time. No assembly of levers, cogs, and pulleys could chat with you about the latest fashions. But the modern computer is different. Since 2007, researchers at IBM have been developing a machine that excels at answering questions put to it in plain English. The researchers call this computer “Watson,” in honor of IBM’s founder, Thomas J. Watson. You can ask Watson anything, on any topic, and it will usually give you the right answer in a few seconds. Watson promises to be the next big step toward creating a machine that speaks and understands a human language. Contrary to what Descartes believed, it now seems inevitable

that a computer will be able to “modify its phrases” to chat with you about clothes, movies, baseball, and anything else.

If this happened—if a computer could respond appropriately to anything you said—would the computer really be intelligent, or would it be just a clever electronic trick? The intelligent robots in science-fiction stories and movies (there are many to choose from) can talk fluently. Should we think of them as conscious beings, with an inner life of thoughts and feelings, or should we think of them only as mechanical contrivances that mimic human sounds and movements?

7.2. An Argument That Machines Could Think

The idea that machines could have conscious thoughts may seem unremarkable to people who have been using sophisticated computers all their lives. Nonetheless, it is a controversial thesis. Here is a line of thought that supports the idea.

The Piecemeal-Replacement Argument. This argument involves a bizarre, unrealistic story. Such stories are called “thought experiments” because you’re merely invited to think about them—you couldn’t possibly set them up in a lab.

The story begins by supposing that a small piece of your brain becomes diseased, and as a result you go blind. But luckily, there is a new biomedical technique that can restore your sight. Scientists have developed a tiny silicon chip, too small to be seen with the naked eye, that duplicates the functions of the diseased part of your brain. So they cut out the diseased bit and replace it with the chip. After the operation, you can see again. Otherwise, you cannot tell any difference. You are as good as new.

A little later, however, a different bit of your brain becomes diseased. This time you lose your ability to remember names. But once again, scientists are able to replace that bit with another tiny chip that does the same job, and when the surgery is over, you can tell no difference except that now you can remember names again.

Now imagine that this happens over and over, until your entire brain has been replaced. At each step in the process, you can detect no difference from before, except that a problem has been corrected. At the end of the process, however, you no longer have an organic brain. Instead, you have an artificial brain, a

patchwork of silicon chips that functions exactly like the original organic model. If this story is possible, then not only could a machine think, but *you* could become such a machine.

There is one further step in the argument. Suppose nothing goes wrong with your brain, and you undergo no operations. Instead, scientists assemble the same parts in the lab and put them into a robot. Would the robot be a conscious, thinking thing? If we grant that this collection of parts could sustain a mental life inside your skull, there is no reason why it should not do the same inside the robot's head. Thus, the robot would be a thinking being in the same sense that you are.

The Objection That Computers Can Do Only What They Are Programmed to Do. At this point a familiar objection may be raised. Computers, it is said, can do only what they are programmed to do. They can never go beyond their programs or display any genuine creativity. A chess-playing computer, for example, just makes moves mechanically, according to its program, without understanding anything about chess. Indeed, when we say it “plays” chess, we are speaking metaphorically—it is really just an electronic device that spits out certain numbers when other numbers are fed into it. It has no idea what chess is. So, there is a big difference between human thinking and computer processing.

This may look like a simple objection, but it raises a host of issues. The objection is partly true, partly hard to assess, and partly irrelevant.

The true part is that present-day chess computers are not conscious of what they are doing. They are relatively simple machines that process inputs and return outputs according to their programs. They have no clue what chess is.

It is harder to say whether present-day computers are genuinely creative. The problem is that we don't know much about what *human* creativity consists in, so it is hard to say whether computers are creative, too. We can certainly point to some of the *results* of human creativity—for example, Einstein's Theory of Relativity, or the invention of Post-it notes, or J. R. R. Tolkien's *The Lord of the Rings*—but we don't understand *how* human beings make creative discoveries. Might human creativity be the result of something like a computer program running in the human brain? If so, then computers could be creative, too. But we just don't know.

Computers *have* given us new and unexpected knowledge in the running of their programs. In this sense, computers can “go beyond” what their programmers put into them. The best-known achievement of a chess-playing computer came in 1997 when Deep Blue, an IBM machine, defeated World Champion Garry Kasparov in a six-game match. This victory of “machine over man” was highly publicized, but the main impact of computers on chess has been to advance chess knowledge: computers have discovered many good moves, which humans have missed, in the analysis of famous chess games, in the analysis of chess openings (during the first 15 or so moves of a chess game), and in the analysis of chess endings (which are chess positions that involve very few pieces).

The limitations of present-day computers, however, are irrelevant to the question of whether computers may someday be conscious. Current chess-playing computers might not be aware of what they’re doing because they only play chess; perhaps we could add some features to them that would make them conscious. To say that this is impossible because a computer will always be executing programs, and nothing more, does not end the debate. Perhaps *we* are only executing programs, but of a very complex kind. Maybe, or maybe not—but the simple assertion that we must be doing something more than executing programs, without saying why, does not advance the discussion.

The Tipping Point Objection. There is a better objection to the Piecemeal-Replacement Argument. The argument assumes that nothing dramatic will happen as your brain is gradually replaced by silicon chips. In particular, it assumes that no single change will cause the lights to go out—no replacement will result in the immediate loss of consciousness.

That assumption can be questioned. Normally, we think of the world as changing gradually: Dust accumulates on the TV set; hairlines recede; riverbanks erode. However, many processes hit thresholds, or *tipping points*: One more drop of water causes the pitcher to spill over; one more banana peel causes the compost heap to catch fire; one more soldier allows the defensive line to be held. Couldn’t a tipping point be reached in the case of piecemeal brain replacement? Perhaps there is some system in the brain that is responsible for consciousness, and when a critical amount of that system has been removed

or replaced, consciousness vanishes. If so, then the Piecemeal Replacement Argument might get its appeal merely from our ignorance—our ignorance of when, in the process of replacement, we would lose consciousness.

7.3. The Turing Test

Most researchers in the field of artificial intelligence (AI) believe that they would be creating a conscious machine if they could program a computer to do everything a human being can do. This belief was held by one of the twentieth century's great figures, the British mathematician Alan Turing.

In 1936, the 24-year-old Turing published a paper called "On Computable Numbers, with an Application to the *Entscheidungsproblem*." The purpose of the paper was to solve a problem in mathematical logic, but in the course of doing so, Turing described, for the first time, the idea of a digital, stored-program, general-purpose computer. At that time, no such machines existed, but Turing understood how they would work.

When World War II broke out, Turing joined the team of code breakers at Bletchley Park, outside London, where he played a key role in deciphering German codes. In the first few years of the war, when the outcome was in doubt, the work at Bletchley was critical. Without it, for example, ships bringing supplies from America could not evade Nazi submarines. To help crack the codes, one of the first electronic computers, the Colossus, was built. Years later, a member of the Bletchley team told an interviewer: "I won't say that what Turing did made us win the war, but I daresay we might have lost it without him." You might expect Turing to have been hailed as a hero after the war. But the work done at Bletchley Park remained top secret for many years, and Turing's contribution remained unknown.

Turing realized that the development of the computer raised a new philosophical question—or at least, it posed an old question in a new, pressing way. In 1950, he published a paper called "Computing Machinery and Intelligence," which began with the words, "I propose to consider the question, Can machines think?" At that time, the question struck most people as absurd, but Turing predicted that, by the end of the century, computers would be so impressive that their "thinking" would no longer seem out of the question. Turing was a smart man.

Turing approached the issue by asking how we might *tell* whether a machine was thinking. The basic idea of the Turing Test is this: Take two humans and a computer, and put them all in different rooms, allowing them to communicate only with typed messages. One of the humans—the interrogator—will ask questions of the other two and try to figure out, from their responses, which is the person and which is the computer. The interrogator can ask anything he likes—he can initiate discussions of poetry, religion, love, or the latest fads. He can try to trip up the computer any way he can. The computer, meanwhile, will deny it is a machine. Turing imagines the following exchange:

Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A: Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R–R8 mate.

Is the responder a human or a machine? The pauses suggest it is a person who needs time for calculating. But the machine might be pausing in order to mimic a human being. After all, it doesn't want to give away that it is a machine.

To make the test fair, we must repeat the experiment several times, with different people involved. After each session, the interrogator guesses which responder is the computer and which is the human being. If no more than half of the guesses are correct (the number that random guessing would produce), then the machine has passed the Turing Test. The computer's responses are indistinguishable from those of a human being.

Turing argued that if a machine passes this test, then the machine must have the mental properties of a person. After all, you take what I say as proof that I am conscious and intelligent. To be consistent, mustn't you also say that a computer giving the *same* verbal performance is equally conscious and intelligent? As one person put it, "If trees could converse with us as fluently

as they do in some fairy tales, wouldn't you unhesitatingly say that trees can think?"

When Turing wrote his original paper, there were only four primitive electronic computers in existence, two in England and two in the United States. Of course, none of them could pass the test. Even today, the Turing Test remains out of reach. Despite the power of present-day computers, none of them can come close to passing it. Indeed, after a half-century of research in artificial intelligence, it is clearer than ever how difficult it would be to create a machine that matches or mimics human intelligence.

One famous AI product, Eliza, can converse in English, but only in a limited way. Eliza is a program developed at MIT that conducts therapeutic interviews with human beings. Here is a conversation between Eliza and a young woman. Eliza's comments are in capital letters:

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR THAT YOU ARE DEPRESSED

It's true that I am unhappy.

DO YOU THINK COMING HERE WILL HELP YOU

NOT TO BE UNHAPPY

I need some help, that much seems certain.

WHAT WOULD IT MEAN TO YOU IF YOU GOT

SOME HELP

Perhaps I could learn to get along with my mother.

TELL ME MORE ABOUT YOUR FAMILY

My mother takes care of me.

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

My father.

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT VERY
AGGRESSIVE

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I AM AFRAID
OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK
OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE
FACT THAT YOUR BOYFRIEND MADE YOU
COME HERE

Mostly, Eliza just turns people's words back on them. This doesn't look very impressive, and Eliza is not hard to trip up—the program can easily be tricked into producing gibberish. Often, however, people are not trying to trip it up. The program's creator, Joseph Weizenbaum (1923–2008), was alarmed at how easily people would come to depend on Eliza, as if she were a human therapist. People would call him demanding Eliza's help. Articles appeared in professional journals suggesting that the program be made more widely available as a diagnostic aid. Meanwhile, Weizenbaum's secretary began to insist that others leave the room so that she could talk to Eliza in private. Weizenbaum himself concluded that AI is dangerous and should be abandoned.

But Eliza lives on. Today, you can talk to her on the Internet. Once, a businessman got connected to Eliza on his company's computer system and had a conversation that he found insulting. He learned only later that he had been talking to a machine. When this happened, it was half-seriously suggested that Eliza had passed the Turing Test. Of course, it had not. Every year, the Loebner Prize is given to the program that comes closest to passing the Turing Test. In recent years, the award has been around \$3,000. If a machine were to actually pass the test, then the award would be \$100,000 plus a solid gold medal. So far, however, no one has come close to claiming the big prize.

Alan Turing did not live to see the advances that his work made possible. In 1952, it was discovered that he was gay. For

this, the man who had broken the Nazi codes was hauled into a British court, stripped of his security clearance, and forced to endure a hormone “treatment” that made him physically grotesque and impotent. In 1954, he committed suicide.

7.4. Why the Turing Test Fails

Is the Turing Test a valid criterion of mentality? Would passing the test prove that a machine was conscious? There are some strong reasons to think not.

First, the test is an application of Behaviorism, a discredited theory of the mind. Behaviorism flourished at mid-century, when Turing did his great work. Behaviorism held that mental concepts such as “thinking” and “understanding” refer to behavior or behavioral dispositions. Turing’s idea was that computers must have certain mental states if they are disposed to respond to human speech in certain ways. But it is now generally accepted that this idea is wrong. Behavior is an expression of mental states, but it is not identical with them, and it does not ensure their presence. A zombie might behave like a person but be mindless. Machines that pass the Turing Test could be such zombies.

Also, John Searle’s Chinese Room Argument seems to show that a computer could pass the Turing Test without understanding anything it said.

The Chinese Room Argument. A man is locked in a room filled with books. In the door is a slot through which slips of paper may be passed. Occasionally, a piece of paper will come through the slot with marks on it, but the marks mean nothing to the man—as far as he is concerned, they are just meaningless squiggles. However, he has been told that, each time a slip of paper comes in, he must send back another slip with different markings on it. The books will tell him what marks to make. The books contain elaborate instructions that say this: Whenever you see a squiggle of such-and-such shape, draw a different squiggle of such-and-such shape. The man does this, hour after hour.

Meanwhile, outside the door, a woman from Beijing is writing messages in Chinese and passing them through the slot. Each time she does this, there is a pause, after which she receives a response in Chinese. The responses make perfect sense, and

in this way she carries on a conversation for several hours. Naturally, she believes that there is someone on the other side of the door who understands Chinese, but she is wrong—there is only the man following the instructions in the books.

We could not create a Chinese Room in the real world because we could not write down all the instructions that would be needed in those books. But suppose we could. What would we think about this situation? Obviously, the man in the room does not understand Chinese. He manipulates symbols according to the instructions he has been given, but he has no way of attaching meaning to those symbols. This seems to show that the Turing Test is unsound. After all, the man in the room passes the Turing Test for understanding Chinese, yet he does not understand Chinese.

Notice that the elements of the Chinese Room correspond to the main parts of a computer: The man is the central processing unit (CPU), the instructions in the books are the program, and the slips of paper are the inputs and outputs. The whole system is doing just what a computer would be doing if a program could be written for conversing in Chinese. The Turing Test says that a computer running such a program would have mental states, including the mental state of understanding its own words. The Chinese Room Argument shows that this is false.

It seems strange that such a simple argument could be so devastating. Yet the many philosophers and cognitive scientists who are skeptical of it have not produced a convincing rebuttal. Perhaps the most popular reply is to suggest that, although the man in the room does not understand Chinese, the whole system of which he is a part does understand it. This is hard to believe. If the man does not understand Chinese, how could the man plus the books understand it? Has the simple interaction between the man and the books somehow generated a new consciousness? Suppose the man memorized all the instructions and discarded the books. The man would then be the system, but he would still not understand Chinese.

The Chinese Room Argument draws its strength from the distinction between *syntax* and *semantics*. Syntactical rules are rules for manipulating symbols. Semantic rules, on the other hand, are rules for interpreting those symbols, for attaching meaning to them. The core of the Chinese Room Argument is this: To have a mind, a system must have more than a syntax. It must also have a semantics. The Turing Test can be satisfied by

a system that has only a syntax. Therefore, the Turing Test is not a valid criterion of mentality.

The “What More Do You Want?” Objection. The Chinese Room Argument has been debated for more than 30 years. Some people are convinced by it, others are not, and some—especially researchers in AI—think it is just silly. Those in the last group are apt to express their frustration like this: If we could create a machine like the ones in the science-fiction stories—a machine that was indistinguishable in its speech and behavior from human beings—then what more could you possibly want? We would have done everything that could be done. We would have as much evidence that the machine thinks as we have that human beings think.

This is an impressive objection. Yet we would *not* have the same evidence that a machine thinks as that a human being thinks, even if the machine passed the Turing Test. To understand a language—to think and to have a semantics—requires consciousness. And our belief that humans are conscious is not merely a deduction from behavior, as it would be for the machine. I know that I am conscious because I am immediately aware of it. Lying in my hammock, I know that I am thinking of the Los Angeles Lakers simply because I am aware of having those thoughts. Moreover, I know that you, as a fellow human being, are in a position like mine. I could not be sure that a computer was in the same position, even if it passed the Turing Test.

So, there is something more that we should want. We should want additional evidence that computers can be conscious. How could we get such evidence? First, we need to understand what it is about the human brain that enables us to be conscious. Right now we know a lot about the brain, but not enough to answer that question in detail. Once we have that information, we can ask whether a computer could have similar features that would enable *it* to be conscious. That would put us on firm ground in thinking about whether machines can think. Unfortunately, that firm ground is not yet available.

I hope we can settle this issue before intelligent robots become a reality. If 200 years from now robots seem to have all the thoughts and feelings of human beings, how to treat them will be a pressing matter. Would they be capable of suffering? Would they have moral rights? Would it be murder to kill them? Technological advances in AI may create new ethical questions, and we’d better be prepared to answer them.

The Case against Free Will

A small part of the universe is contained within the skin of each of us. There is no reason why it should have any special physical status because it lies within this boundary.

—B. F. SKINNER, *ABOUT BEHAVIORISM* (1974)

8.1. Are People Responsible for What They Do?

In 1924, two Chicago teenagers, Richard Loeb and Nathan Leopold, kidnapped and murdered a boy named Bobby Franks just to prove they could do it. The crime caused a sensation. Despite the brutality of what they had done, Leopold and Loeb did not appear to be especially wicked. They came from rich families and were both outstanding students. At 18, Leopold was the youngest graduate in the history of the University of Chicago, and at 19, Loeb was the youngest graduate ever from the University of Michigan. Leopold was about to enroll at Harvard Law School. How could they have committed a senseless murder? Their trial would receive the same level of attention that the O. J. Simpson trial did 70 years later.

The parents hired Clarence Darrow, the most famous lawyer of the day, to defend them. Darrow was a champion of unpopular causes—he had defended labor organizers, communists, and a black man accused of killing a member of a racist mob. In 1925, in his most celebrated case, he would defend John Scopes of Tennessee from the charge that he had taught evolution in a high school classroom. Darrow was also the country's best-known opponent of the death penalty. In 1902, he had

been invited by the warden to give a talk to the inmates of the Cook County Jail in Chicago, and he told the prisoners:

I really do not in the least believe in crime. There is no such thing as a crime as the word is generally understood. I do not believe there is any sort of distinction between the real moral conditions of the people in and out of jail. One is just as good as the other. The people here can no more help being here than the people outside can avoid being outside. I do not believe that people are in jail because they deserve to be. They are in jail simply because they cannot avoid it on account of circumstances which are entirely beyond their control and for which they are in no way responsible.

These ideas would figure prominently in Darrow's defense of Leopold and Loeb.

The public wanted blood. As the trial began, the *Chicago Evening Standard* carried this headline:

DARROW PLEADS FOR MERCY: MOBS RIOT

BAILIFF'S ARM BROKEN AND WOMAN FAINTS AS FRENZIED MOB
STORMS PAST GUARDS; JUDGE CALLS FOR 20 POLICE; FEARS SOME
WILL BE KILLED

Leopold and Loeb had already admitted their guilt, so Darrow's job was just to keep them from the gallows. There would be no jury. The judge would listen to the lawyers' arguments and then decide whether the defendants would hang.

Darrow spoke for more than 12 hours. He did not argue that the boys were insane; nevertheless, he said, they were not responsible for what they had done. Darrow appealed to a new idea that psychologists had proposed, namely, that human character is shaped by an individual's genes and environment. He told the judge, "Intelligent people now know that every human being is the product of the endless heredity back of him and the infinite environment around him."

I do not know what it was that made these boys do this mad act, but I do know there is a reason for it. I know they did not beget themselves. I know that any one of an infinite number of causes reaching back to the beginning might be working out in these boys' minds, whom you are asked to hang in malice and in hatred and injustice, because someone in the past sinned against them.

Psychiatrists had testified that the boys lacked normal feelings, because they showed no emotional reaction to what they had done. Today a psychiatrist might say that Leopold and Loeb suffered from “antisocial personality disorder,” commonly known as “sociopathy” or “psychopathy.” Darrow seized upon this idea:

Is Dickie Loeb to blame because out of the infinite forces that conspired to form him, the infinite forces that were at work producing him ages before he was born, that . . . he was born without [the right kind of emotions]? If he is, then there should be a new definition for justice. Is he to blame for what he did not have and never had?

Darrow portrays Loeb as having had a childhood bereft of the affection that boys need, spending his days studying and his evenings secretly reading crime stories, fantasizing about committing the perfect crime and fooling the cops. Leopold, meanwhile, was weak and without friends. He grew up obsessed with Nietzsche’s philosophy of the “superman,” disdaining other people and desperately wanting to prove his own superiority. Then the two boys found one another and committed the crime. But they were just playing out the hand nature dealt them. “Nature is strong and she is pitiless,” Darrow concluded. “She works in her own mysterious way, and we are her victims. We have not much to do with it ourselves.”

The judge deliberated for a month and then sentenced Leopold and Loeb to life in prison. Twelve years later, Richard Loeb was attacked and killed by another prisoner. For his part, Nathan Leopold spent 34 years behind bars. During that time, he taught other prisoners, volunteered for malaria testing, ran the prison library, and worked in the prison hospital. After his release on parole, he moved to Puerto Rico, where he continued his lifelong effort to “become a human being again,” largely through jobs that involved helping others. He died in 1971.

8.2. Determinism

Clarence Darrow was the first lawyer to use the defense that people are never responsible for their actions, because their actions are caused by forces beyond their control. However, Darrow was not the first person to doubt that we choose our own destinies.

Aristotle worried that the laws of logic might imply that we have no choice about what we do. Every proposition, he reasoned, must be either true or false. So at this moment it is either true or false that you will drink a Diet Coke tomorrow. If it is true, then there is nothing you can do to prevent it—after all, *it will happen*. If it is false, then there is nothing you can do to make it happen, for *it will not happen*. Either way, the future is fixed, and you cannot change it. This became known as the problem of Fatalism. Theologians from St. Augustine onward realized that the assumption of God's omniscience creates a similar difficulty. If God knows everything, he knows what you will do tomorrow. But if God already knows what you will do, then you cannot do otherwise.

Fatalism is a serious problem, but it is not the biggest challenge to human freedom. A greater threat is posed by Determinism, which was known in the ancient world but came into its own with the rise of modern science. To say that a system is deterministic means that everything that happens in it stems from prior causes. Once the causes occur, the effects must follow, given the surrounding circumstances and the Laws of Nature. For example, you expect the electrical devices in your home to be deterministic systems. If your reading lamp goes out, you expect there to have been a cause: the filament in the bulb burned out, or your little brother pulled the plug out of the socket, or something. If you could not fix the problem, and the repairman said, "It just happened, for no reason," that would violate your conception of how things work.

With the rise of modern science, it became common to think of the whole universe as one giant deterministic system. Nature consists of particles that obey the laws of physics, and everything that happens is governed by the unchanging laws of cause and effect. This idea was vividly expressed by the French mathematician Pierre-Simon Laplace (1749–1827), who said that if a supremely intelligent observer knew the exact location and velocity of every particle in the universe and all the laws of physics, he could predict with certainty every future state of the universe. Nothing would surprise him; he would know everything before it happened. As Laplace knew, we cannot actually make such predictions, but only because we lack the necessary information.

The universe includes us. We are part of nature, and what happens inside our skins is subject to the same physical laws as everything else. The movements of our arms, legs, and tongues are triggered by events in our brains, which in turn are caused by other physical occurrences. Thus, Laplace's perfect observer could predict our actions in the same way that he predicts everything else. In fact, by tracing the causes far enough back, he could have predicted whether you will drink a Diet Coke tomorrow even before you were born. It may *seem* to us that we make our choices freely and spontaneously, but Laplace argued that our "freedom" is only an illusion created by our ignorance. Because we aren't aware of what causes our behavior, we assume that we act freely.

What, exactly, are the underlying causes of our behavior? As Darrow observed, the "ultimate" causes may stretch far back in time. But the immediate causes are events in our brains. Neurological events cause both our mental states to occur and our bodies to move. This last claim is not mere speculation. Brain surgery sometimes takes place under only a local anesthetic, so the patient can tell the surgeon what he or she is experiencing as various parts of the brain are probed. This technique was pioneered more than a half-century ago by Dr. Wilder Penfield, who described it in his book *The Excitable Cortex in Conscious Man* (1958). Neurosurgeons have been using Penfield's technique ever since. They know that if you probe in one place, the patient will feel a tingle in her hand; and if you probe in another place, the patient will smell garlic. In one patient, probing in still another place made her hear a song by Guns n' Roses.

Actions can also be induced by electrical stimulation of the brain. Jose Delgado (1915–), a Yale University scientist, discovered that stimulating various regions of the brain could cause all sorts of bodily motions, including frowning, the opening and closing of the eyes, and movements of the head, arms, legs, and fingers. When he first tried this using cats and monkeys, Delgado noticed that the animals showed no surprise or fear when their bodies moved. Apparently, they experienced the movements as voluntary. In one instance, stimulating a monkey's brain caused the monkey to get up and walk around. The effect was repeated several times, and each time the animal strolled around without surprise or discomfort, as if it had just decided to take a walk.

Some philosophers would say that Delgado's technique produced only bodily movements, not actions, because actions involve reasons and decisions. But there is more. When Delgado tried his experiment on humans, not only did they act out the movements without surprise or fear, but they also produced reasons for them. In one subject, electrical stimulation of the brain produced "head turning and slow displacement of the body to either side with a well-oriented and apparently normal sequence, as if the patient were looking for something." This was repeated six times over two days, confirming that the stimulation was actually producing the behavior. But the subject, who did not know about the stimulation, considered the activity spontaneous and offered reasons for it. When asked "What are you doing?" he would reply, "I am looking for my slippers," "I heard a noise," "I am restless," or "I was looking under the bed."

Are our *decisions* also produced by neural firings? There are some experimental results about this, too, due to the German scientist H. H. Kornhuber. Suppose you sit quietly, and sometime during the next minute you spontaneously move your finger. Subjectively, you may feel certain that the decision to move your finger is entirely within your control. But now suppose we attach some electrodes to your scalp and ask you to repeat the action. A technician watching an EEG (an electroencephalograph) would be able to observe a characteristic pattern of brain activity when you move your finger. The brain activity begins up to one-and-a-half seconds before the movement, and *it begins before you make your decision*. So the technician, watching his monitor, knows that you are going to move your finger before you do. He is, in a small way, like Laplace's perfect observer. Kornhuber first performed this experiment in the 1970s.

8.3. Psychology

It may seem odd that the primary argument against free will appeals to the principles of physics. After all, psychologists, not physicists, study human behavior. So we might wonder what psychology has to say. Do psychological theories about human behavior have room for the notion of free will, or do they support Determinism?

Before turning to psychology, however, let me mention some of the ways in which our commonsense understanding of human beings already contains elements favorable to Determinism. Each of us was born to particular parents at a particular time and place, and only a little thought is needed to realize that if those circumstances had been different, we would have turned out different. A young man “chooses” to become a stockbroker—is it a coincidence that his father is a stockbroker? What would he choose if his parents were missionaries? What would he have chosen if he had been born into a different culture?

We also know from statistics that social conditions influence us in ways we don’t realize. Consider, for example, the names we give to our children. White parents tend to choose different names than black parents. In California in the 1990s, for example, girls named Molly and Amy were overwhelmingly white, while girls named Imani and Ebony were overwhelmingly African-American. Social differences between the white and black communities resulted in different names being popular. Socioeconomic status also influences our naming decisions. Again in California, the most common names given to rich white girls in the 1990s were Alexandra, Lauren, Katherine, and Madison; poor white girls were most often called Amber, Heather, Kayla, and Stephanie. These trends change over time. Sometimes “rich names” become more popular among the poor, which in turn causes the rich to abandon them. Individually, parents may always seem to be making free, independent choices. But when social circumstances change, so does the popularity of the names.

Or consider another example: The United States has the highest rate of imprisonment in the world. Over two million Americans are now in jail or prison, and five million more are on probation or parole. Here are some social factors that have been shown, statistically, to be predictive of criminal behavior: having an uneducated mother, growing up poor, being raised in a single-parent household, and having a teenage mother. Considered as individuals, it may seem that each person “freely decides” to break the law. However, many of these individuals would not have broken the law, had their circumstances been different.

When we set aside statistics and try to understand in more detail why individuals behave as they do, we always seem to wind

up with explanations in which “free choice” plays little part. Darwin’s explanation of how Leopold and Loeb came to kill Bobby Franks is one example. For another, consider Eric Rudolph, who committed a series of bombings in Georgia and Alabama in the 1990s. I heard one of the explosions from my office—the bomb killed a policeman and critically wounded a nurse at an abortion clinic in Birmingham, Alabama. Rudolph was caught in 2003, and today he is serving five consecutive life sentences.

Why would he have done such a thing? Why did he kill the policeman and maim the nurse in Alabama? We might suppose that Rudolph hated abortion so much that he was willing to use any means to stop it. That may be true, but a lot of people oppose abortion without planting bombs. Why did this particular man turn deadly?

According to *Newsweek* magazine, “He is perhaps best understood as the product of a paranoid fringe of white supremacists, religious zealots and government haters. Rudolph’s mind and motives are hard to fathom, but extremism seems to run in the family.” When he was 13, Eric’s father died, and his family moved from Miami to rural North Carolina. There they lived on a gravel road near a sawmill owner named Tom Branham. Branham, a survivalist who had been arrested on federal weapons charges, took an interest in Eric and his brother Daniel and became a substitute father to the boys. His mother, meanwhile, moved to the Missouri Ozarks to join a community of white separatists. By the time he was in the ninth grade, Eric was writing in a class paper that the Holocaust never happened, using as his “research” pamphlets issued by hate groups. As much as we might detest what Eric Rudolph became, it is hard to believe that the little boy had a chance.

Classical Psychology. The effort to understand human behavior naturally leads us to consider deterministic explanations. “He just decided to do it” doesn’t sound like any explanation at all. Thus, Determinism is a hard-to-avoid by-product of the search to explain behavior. As B. F. Skinner (1904–1990), a Harvard University psychologist, put it:

If we are to use the methods of science in the field of human affairs, we must assume that behavior is lawful and determined. We must expect to discover that what a man does is the result of specifiable conditions and that once

these conditions have been discovered, we can anticipate and to some extent determine his actions.

As the science of psychology has developed, one theory after another has competed for acceptance. But none of them has had much use for “free will.”

For most of the twentieth century, Behaviorism and Freudianism were the most popular theories among psychologists. Freud sought to understand human conduct by identifying the unconscious motives of action. Conscious processes of thought and deliberation are, on his view, only rationalizations for deeper forces hidden within the psyche. Long-forgotten events of infancy and early childhood have given each of us unconscious desires and impulses that control us as adults. For example, a woman has a series of relationships with abusive men. Each time she gets rid of one, she vows never to make that mistake again; but she does, over and over. How can she keep doing this? She appears to choose freely each time she begins a new relationship, but really she does not. She has a masochistic personality, formed as a child when she was battered by an abusive father; now, as an adult, she helplessly reenacts her relationship with her father again and again. She will not be able to break the pattern until she confronts her repressed memories and feelings about her father, possibly after years of psychoanalysis.

The behaviorists would have none of this. On their view, unconscious thoughts play no part in explaining behavior. In fact, no thoughts of any kind do. Instead, a person's behavior is explained by the process of conditioning that produced that behavior. We tend to repeat behavior for which we are rewarded, and we tend not to repeat behavior when we are punished. Suppose you get an electric shock every time you touch a fence; you will soon stop touching it. Or suppose that a child is fed when he says “please” and not fed when he says nothing. He will soon be saying “please” whenever he is hungry. These are simple examples, but the principle is supposed to be the same for all behavior. In theory, everything we do can be explained as a response to prior conditioning, including our proudest and noblest actions as well as our most shameful ones.

Is “Character” a Matter of Luck? When we reflect on other people's misfortunes, we sometimes think, “There but for the grace of God go I.” Are most of the moral differences between

people a matter of luck? Some of the most famous experiments in social psychology suggest that any of us might behave badly if we were unlucky enough to be in the wrong circumstances.

In one experiment, Philip Zimbardo and his colleagues set up a simulated prison in the basement of a Stanford University building. Twenty-four volunteers were assigned to be either guards or prisoners. The experiment was supposed to last two weeks, but it had to be called off after five days because the “guards” treated the “prisoners” so brutally.

In another study, Stanley Milgram asked volunteers to operate a device that administered increasingly severe electric shocks to someone in the next room. The person in the other room was supposed to be “learning” by being punished for giving the wrong answers to questions. (He was, in fact, an actor who was only pretending to be shocked.) Milgram was surprised to discover that every single volunteer was willing to continue shocking the other person even when the levels were labeled as extremely dangerous and the other person could be heard crying and begging the volunteer to stop.

When people hear about these experiments, they invariably feel that *they* would not have acted so badly. This feeling is hard to avoid, yet the participants in the experiments were ordinary people like you and me. Zimbardo comments that, after the guards-and-prisoners experiment was over, the “prisoners” insisted that they would not have been so abusive if they had been guards. But there was no difference between those who were made guards and those who were made prisoners—the assignment was random. The natural conclusion is that the only difference between them was in their circumstances. Apparently, all of us—or at least the great majority of us—have the inner capacity to behave badly if we are in the wrong circumstances.

One other study deserves mention: the “Good Samaritan” experiment of J. M. Darley and C. D. Batson. In Luke’s gospel, the Good Samaritan is presented as a model of decent behavior:

“And who is my neighbor?”

Jesus replied, “A man was going down from Jerusalem to Jericho, and he fell among robbers, who stripped him and beat him, and departed, leaving him half dead. Now by chance a priest was going down the road; and when he saw him, he passed by on the other side. So likewise a

Levite, when he came to the place and saw him, passed by on the other side.

“But a Samaritan, as he journeyed, came to where he was; and when he saw him, he had compassion and went to him and bound his wounds, pouring on oil and wine; then he set him on his own beast and brought him to an inn, and took care of him. And the next day he took out two denarii and gave them to the innkeeper, saying, ‘Take care of him; and whatever more you spend, I will repay you when I come back.’ Which of these three, do you think, proved neighbor to him who fell among the robbers?”

He said, “The one who showed mercy on him.”

And Jesus said to him, “Go and do likewise.”

The traditional interpretation of this story is that the Samaritan was a man of better moral character—he “had compassion,” while the priest and the Levite did not. (Samaritans, incidentally, were people of low standing, while priests and Levites played important roles in the Temple.) Is this right? Darley and Batson decided to investigate the circumstances in which we would be Good Samaritans, using Princeton theology students as their subjects.

In the study, the theology students first filled out forms giving information about themselves, including information about their ethical and religious beliefs. Then half of the students were told to prepare a lecture on ethics, and half were told to prepare a lecture on job opportunities. All the students were told to go to another building to give their lectures. Some were told they needed to hurry, while others were told they had plenty of time. It had been arranged that on their way to the other building they would pass by someone slumped in a doorway, obviously in distress. Would they stop to help?

Some stopped and some did not. But it turned out that their ethical and religious views had nothing to do with it, nor did it matter whether they had ethics or job opportunities on their minds. All that mattered was whether they thought they had time to stop. This small change in circumstances determined who would be heroic and who would be heartless.

8.4. Genes and Behavior

There aren’t many Freudians or behaviorists around these days. Both theories try to do too much with too little. One limitation is that, by focusing only on childhood and environmental

conditioning, neither theory has much use for genetic explanations. Today, researchers believe that our genes are about as important as the environment in shaping our personalities.

Twin Studies. How can we estimate the influence of genes on human behavior? One strategy is to compare identical twins who were separated at birth and raised in different environments. Typically, these are twins who were adopted by different families. Dissimilarities between the twins, as adults, cannot be due to genetic differences—after all, identical twins have the same genes. Similarities between the twins might be due either to genes or to common environmental factors (even different households will have some things in common). Twin studies can give us at least a rough idea of how much influence genes exert.

At the University of Minnesota, there is an ongoing research project, begun in 1979, called the Minnesota Study of Twins Reared Apart. When such twins are located, they are invited to the university for a series of tests. The researchers have found that identical twins reared apart resemble each other very closely. In some cases, the similarities sound like something out of science fiction.

Among those tested were the “Giggle Sisters,” both of whom laughed all the time. Both had the habit of pushing at their noses, which they both called “squidging”; both claimed to have weak ankles as a result of falling when they were 15; both had met their husbands at dances when they were 16; and both had worked as polling clerks.

There were also brothers named Jim, who drove the same model car and smoked the same brand of cigarettes. Both had elaborate workshops at home where they made miniature furniture as a hobby. Both liked to leave little love notes for their wives lying around the house. And they had named their sons James Alan and James Allan.

Most remarkable, perhaps, were Jack Yufe and Oskar Stöhr, whose home environments had been exceedingly different. One twin was raised in Trinidad by his Jewish father, the other in Germany by his Nazi grandmother. Oskar was in the Hitler youth; Jack served in the Israeli navy. When reunited, both were wearing rectangular wire-frame glasses and blue two-pocket shirts with epaulets. Both had small mustaches. Both liked to read magazines from back to front, and

both flushed toilets before using them. Also, both men liked the same odd practical joke of startling people by sneezing in elevators.

These are great anecdotes, but we can draw no firm conclusions from them. In the first place, an enormous amount of data would have to be gathered and analyzed before we could know what, if anything, to conclude. Take the blue-shirt-with-epaulets story, for example. How many shirts like that were owned by men in the areas where Jack and Oskar lived? What are the chances of two men from those areas wearing that shirt on the same day? Or, more generally, in any group of males drawn from the same population, what are the chances that any two selected at random will be dressed similarly? Most important of all, what are the chances that there will be *some* striking similarity between two such men, even if it is not how they are dressed? Critics have also said that the anecdotes might be exaggerated. Moreover, it turns out that some of the twins had met one another before they were studied in Minnesota. Thus, some of the similarities might have grown out of their interactions.

The researchers do not, however, base their conclusions on such anecdotes. Instead, the reared-apart twins are given standard psychological tests for such traits as flexibility, tolerance, conformity, self-control, conscientiousness, openness, tough-mindedness, social dominance, alienation, authoritarianism, and aggressiveness. And the twins have been found to be remarkably alike in all these ways. They have similar senses of humor and similar levels of optimism or fearfulness. They share (or lack) similar talents, and they have similar mental illnesses and disabilities. On the basis of such studies, researchers have concluded that the major components of our personalities are about 50 percent due to our genes. And our personalities, of course, affect our behavior.

Are Some People Born Bad? Scientists have long known about the connection between biology and violence. Experiments have shown that if a small section of a cat's hypothalamus is removed, the cat will turn savage. Humans with head injuries sometimes experience fits of uncontrollable rage. Meanwhile, for people who are "naturally" prone to violence, effective treatments include lithium and beta-blockers, which

act on the brain. The genes–neurology–violence connection was further confirmed in 1995 when geneticists discovered that turning off the gene responsible for producing nitric oxide—a chemical in the brains of both mice and humans—causes normally sociable mice to become vicious. So the fact that there is *some* sort of connection between genes, neurology, and violence is well established.

Some researchers believe that this tells us something important about crime. The general notion of “crime” is too socially variable to be of much use—fornication, gambling, and heresy, for example, are sometimes counted as “crimes” and sometimes not. But suppose we focus on violent crimes—for example, murder, assault, and rape. Is it “in the genes” for some people, but not others, to do violence? The evidence suggests so.

Darrow believed that Leopold and Loeb were “born bad” because they were born without such feelings as pity and sympathy. There is no way for us to know the precise truth about Leopold and Loeb, but on the more general issue Darrow might have been right. The psychologist Judith Rich Harris (1938–) puts it like this:

Though we no longer say that some children are born bad, the facts are such, unfortunately, that a euphemism is needed. Now psychologists say that some children are born with “difficult” temperaments—difficult for their parents to rear, difficult to socialize. I can list for you some of the things that make a child difficult to rear and difficult to socialize: a tendency to be active, impulsive, aggressive, and quick to anger; a tendency to get bored with routine activities and to seek excitement; a tendency to be unafraid of getting hurt; an insensitivity to the feelings of others; and, more often than not, a muscular build and an IQ a little lower than average. All of these characteristics have a significant genetic component.

It is easy to understand why such ideas arouse controversy. It sounds like we are being told that some children are beyond help—they’re born bad, and bad they’ll stay. But the fact that genes contribute to some undesirable traits does not imply that the children who have them are lost causes. For one thing, researchers now believe that there are “orchid children” who have the genetic tendency to “wilt if ignored or maltreated but bloom spectacularly with greenhouse care.” If such children appear to be beyond help, this is only because they haven’t

yet been placed in the right environment. Moreover, no social scientist believes that genes determine everything. Your genes might incline you, in certain environments, to act in certain ways, but whether you actually behave in those ways will depend on other things. Thus, education and the elimination of poverty are still important. The research about genes only explains why virtue comes easier to some than to others.

But even if ill-behaved children might improve, we are being told something disturbing: Human beings come equipped by nature with deep-seated desires that we can resist only with difficulty. If, in some people, these desires prove irresistible, it is hard to see this as their fault. Moreover, these desires might be with us forever, and they play a significant role in our behavior. This may not be Determinism, but it sounds suspiciously close to it.

Conclusion. The case against free will comes from several parts of science. In general, scientists assume that the universe is governed by the laws of nature—by the laws of physics, chemistry, and biology. So far, everything we know about the human brain suggests that it is no exception: neurological events, including the events that cause us to act, appear to be governed by the same natural laws that govern everything else. Meanwhile, psychologists have developed a number of ideas that help explain human action. We don't know how to combine these ideas into a single, comprehensive theory of human nature, but the overall trend is not encouraging to believers in free will. Each new discovery chips away a bit more of our confidence. The more we learn about the sources of human conduct, the less room there seems to be for free choice.

The Debate over Free Will

Man's life is a line that nature commands him to describe on the surface of the earth, without his being able to swerve from it, even for an instant. . . . Nevertheless, in spite of the shackles by which he is bound, it is pretended he is a free agent.

—PAUL HENRI BARON D'HOLBACH,
THE SYSTEM OF NATURE (1770)

Sir, we *know* our will is free, and *there's* an end on it.

—DR. JOHNSON, IN BOSWELL'S
THE LIFE OF SAMUEL JOHNSON, LL.D. (1791)

9.1. The Determinist Argument

When asked about his philosophy of life, the novelist Isaac Bashevis Singer (1904–1991) liked to tell his friends: “I believe in free will. I have no choice.” Singer’s little joke makes a serious point: It is hard to avoid thinking that you have free will. When you are deciding what to do, the choice seems entirely yours. The inner feeling of freedom is so powerful that we may be unable to give up the idea of free will, no matter how much evidence there is against it.

And there is a lot of evidence against it. The more we learn about the causes of human behavior, the less likely it seems that we’re free. No one piece of evidence forces this conclusion. Nonetheless, many different strands of evidence point in this direction, and the cumulative effect is that “free will” looks more and more like part of a prescientific way of thinking.

According to Determinism, everything that happens must happen, given the Laws of Nature and the past history of the universe. We may call this the Determinist Argument:

- (1) Everything we do is caused by forces over which we have no control.
- (2) If our actions are caused by forces over which we have no control, we do not act freely.
- (3) Therefore, we never act freely.

This is a disturbing line of thought because of what it seems to imply about individual responsibility. If we are not free, then it seems that we are not responsible for what we do.

Why should we accept the first premise—why should we believe that our actions are caused by forces outside of our control? The main reason is that we are part of nature. Our bodies and our brains are composed of atoms, and like everything else composed of atoms, they are subject to the laws of physics. And we, of course, have no control over the laws of physics—these laws existed before we were born, and they will continue to operate after we die. Therefore, human actions are just like other natural events—they are due to the Laws of Nature and to prior circumstances over which we have no control.

Is the Determinist Argument sound? It is plausible, but it can be disputed. We will look at two responses to the argument. One theory, Libertarianism, denies the first premise of the argument and holds that our actions are not determined by forces beyond our control. Another theory, Compatibilism, denies the second premise and holds that we can be free *even if* our actions are determined by outside forces.

9.2. The Libertarian Response

Libertarianism is the view that at least some of our actions are not forced on us by the Laws of Nature. Rather, we freely choose to perform those actions, and nothing makes us do so. We could act differently, but we choose not to. According to this theory, human choices are not constrained like other events are. A billiard ball, when struck by another billiard ball, must move in a certain direction with a certain velocity. It has

no choice; the laws of physics determine precisely what will happen. But right now, you can decide whether to continue reading or to stop reading. You can do either, and nothing makes you choose one way or the other. The Laws of Nature do not control you.

Several arguments have been given in favor of this theory.

The Argument from Experience. We may begin with the idea that we know we are free because *each of us is immediately aware of being free* every time we make a conscious choice. Think again of what you are doing at this moment. You can continue reading, or you can stop. Which will it be? Think of what it is like, right now, as you consider these options. You feel no constraints. Nothing is holding you back or forcing you in either direction. The decision is up to you. The experience of freedom, it may be said, is the best proof we could have on the matter. As Samuel Johnson told his friend Boswell, “You are surer that you can lift up your finger or not as you please, than you are of any conclusion from a deduction of reasoning.”

The problem is that the evidence against free will undermines our confidence in our experience. The scientist Jose Delgado found that he could cause people to do things by electrically stimulating their brains—he could get them to look over their shoulder, for example—and then they would offer reasons why they had done so. They would say things like, “I was looking for my comb.” Delgado’s subjects, who did not know their brains were being stimulated, would *experience* the movements as free and voluntary. Perhaps we are like Delgado’s subjects. What is the difference between us and them, except that in the experimental setting we know about the electrochemical events that cause the action, while in everyday life we don’t know what’s happening in our brains? In the face of such evidence, it is hollow to insist, with Dr. Johnson, that we simply know we are free. If freedom is to be defended, a better argument is needed.

The Argument That the Universe Is Not a Deterministic System. Determinism goes against present-day science. In the heyday of Newtonian physics, the universe was believed to operate according to strict laws of cause and effect. The Laws of Nature were believed to specify the conditions under which one state of affairs *must* follow another—the motion of balls on a pool table

was a model for the whole universe. But in the twentieth century, Newtonian physics was replaced by a theory called quantum mechanics, which implied a different view of nature.

According to quantum mechanics, the laws that govern the behavior of matter are probabilistic. The laws do not say, "Given X, Y *must* follow." Instead, they say, "Given X, there is *a specific probability* that Y will follow." For example, suppose that a radioactive substance is decaying. The Laws of Nature may tell us that a certain percentage of the atoms are likely to decay in the next month. However, the laws will not tell us exactly which atoms will decay. That's left undetermined.

Some scientists believe, on philosophical grounds, that quantum mechanics cannot be right. "God does not play dice with the universe," said Albert Einstein (1879–1955). Einstein believed that quantum mechanics was incomplete. Eventually, he thought, we would learn more, and our new theory would be deterministic. So far, however, Einstein's prediction has proved wrong. Quantum mechanics, with its built-in uncertainties, remains one of the most powerful and successful theories in the history of science.

Does this mean that we needn't worry about Determinism? Quantum mechanics has sometimes been hailed as good news for free will. If some things are not determined by the Laws of Nature, it is said, then we might be free, because our actions might be among those undetermined things.

However, quantum physics does not offer much help in defending free will, for two reasons. First, quantum mechanics interprets nature as containing a lot of randomness—it is random and unpredictable which radioactive atoms will decay. However, these random events occur everywhere in nature, not just in human bodies. Is free will supposed to pervade every bit of matter, in every corner of the universe? No, because freedom is not the same thing as randomness. Second, the implications of quantum indeterminacy for human behavior are probably very small. Consider the implications of quantum mechanics for computers. A computer's outputs are determined by its inputs and its program. Quantum mechanics does not imply that we should stop trusting computers. According to quantum theory, a computer might randomly deviate from its program, but the chances of this happening are very, very small. The computer's operations are, if not

completely determined, then close enough as to make no difference. We will almost always get the expected outputs, given the right program and the right inputs. Something similar could be true of human beings. If so, that would be enough for the Determinist Argument.

The Argument That We Cannot Predict Our Own Decisions.

A different argument for Libertarianism exploits the idea that anything which is causally determined is predictable. The tree outside my window is leaning, and sooner or later it will fall. If I knew all the Laws of Nature, and if I knew everything about the tree and its physical environment, I could predict exactly when it will fall. Of course, I don't know those things, so I can't make that prediction. We might say that causally determined events are predictable *in principle*, if not always in practice.

If human actions are determined by forces outside our control, then they too should be predictable in principle. We would only need to know the pertinent facts about the person, the person's circumstances, and the relevant Laws of Nature. However, you could not always predict your own actions, even in principle. For example, suppose you were trying to decide between going out for pizza with your friends and staying home and having leftovers. Could you, in principle, predict what you'll do? Even if you knew everything about your brain, you couldn't do it. The problem is that, once you made the prediction, you could change your mind, just to keep things interesting. If you predict that you'll go out with your friends, then you could decide to stay home, just to prove a point about free will. This seems to show that there is a big difference between human actions and other events.

We may summarize the argument like this:

- (1) If human actions are determined by forces beyond our control, then they are in principle predictable.
- (2) But a prediction about human action can be thwarted if the person knows about the prediction and chooses to act differently.
- (3) Therefore, not all human actions are in principle predictable.
- (4) And so, not all human actions are determined by forces beyond our control.

This is a clever argument, but is it sound? Unfortunately, the argument fails once we examine what the word “predictable” means. There are two types of predictability:

- (a) Predictable by an ideal observer who stands outside the system and observes events but does not interfere with them
- (b) Predictable by you in the real world

Determinism implies predictability in sense (a) but not in sense (b). Suppose you predict you’ll have pizza, and then you decide against it. Your prediction was indeed proved wrong. However, an ideal observer might have known exactly what was going to happen—she might have been able to predict your prediction as well as your decision to do the opposite. Everything might still be determined.

This argument, then, doesn’t prove that our actions are up to us. But there is one more argument to consider.

The Argument from Accountability. The assumption that we have free will is deeply engrained in our ordinary ways of thinking. As we interact with other people, we cannot help but think of them as the authors of their actions. We hold them responsible, blaming them if they behave badly and admiring them if they behave well. For these attitudes to be justified, people must have free will.

Similarly, in thinking about ourselves, we assume that we have free will. Someone who hits the winning basket or aces a test may feel proud, while someone who chokes or cheats may feel ashamed. These feelings of pride and shame would be baseless if our actions were always due to factors beyond our control. Yet these feelings are inescapable. We must think of human beings as free.

We might therefore reason as follows:

- (1) We cannot help but admire or blame people for what they do, nor can we avoid feeling pride or shame for what we do.
- (2) These responses—admiration, blame, pride, and shame—presuppose that people have free will.
- (3) Therefore, we must believe that people have free will.
- (4) Since we must believe it, it is true: People have free will.

The problem with this argument is obvious: Our attitudes might be unjustified, even if they're inescapable. Step 4 is therefore unwarranted. The argument engages in a kind of wishful thinking.

Is Libertarianism Coherent? Finally, we may consider whether Libertarianism makes any sense as a positive view of human behavior. To understand our behavior, we need more than just the denial that our actions are determined by outside forces. We need, in addition, a positive account of how we make decisions.

If our actions are not determined by the Laws of Nature, how are they supposed to come about? What, exactly, produces our decisions? We might imagine that there is, inside each of us, a sort of "mental being" whose decisions are not constrained by the Laws of Nature—a ghostly controller who makes choices independently of the happenings in the brain. But this is not credible. It goes against what science tells us. There is no evidence for any sort of "mental energy" at work within us, disconnected from the operation of our neurological systems.

But if we don't say that a disconnected mental entity inside us calls the shots, what do we say? Are we supposed to think that some part of the brain operates outside the causal network of the world? It sounds silly, but it is hard to come up with anything better. There seems to be no plausible story that makes sense of Libertarian "freedom." Without such an account, we must look elsewhere for a solution to the problem of free will.

9.3. The Compatibilist Response

Compatibilism is the idea that an act can be both free and determined at the same time. This may sound like a contradiction, but according to this theory, it is not. Contrary to what you might think, we can accept the idea that human behavior is free while also acknowledging that it is determined by the Laws of Nature.

Compatibilism has always enjoyed a following in philosophy. In one form or another, it was the theory of Hobbes, Hume, Kant, and Mill, and it is defended by many writers today. This usually comes as a surprise to people who are not familiar with the philosophical literature, because free will and Determinism seem obviously *incompatible*. How are they supposed to go together? How can an act be free and determined at the same time?

According to Compatibilism, some actions are obviously free, and some are obviously not free. The trick is to see the difference between them. Here are some examples of when your action is not free:

- You hand over your wallet because a robber points his gun at you.
- You attend the company picnic because your boss tells you to.
- You report for induction into the army because you've been drafted.

In these cases, you are not acting freely because you are being forced to do things you don't want to do. On the other hand, here are some cases in which you do act freely:

- You contribute money to a charity because you believe that the charity deserves your support.
- You urge your company to sponsor a picnic because you think it would be fun. You're delighted when your boss agrees, and you volunteer to organize the event.
- You join the army because you want to fight the enemy.

These actions are free because your choice is based on your own desires, without anyone else telling you what to do. *This is what it means to do something "of your own free will."* But notice that this is perfectly compatible with your actions being causally determined by your past history, by events in your brain, and so on. It is even compatible with your *desires* being caused by factors beyond your control. The theory says that, to act freely, you must act according to your desires. However, the theory says nothing about where those desires came from. Thus, free will and Determinism are compatible.

According to Compatibilism, a free action might be predictable. Is that a problem for the theory? Consider an example. I have a friend who sees lots of movies, and I know what sorts of movies she likes. I have been observing her moviegoing habits for years. If she is picking a movie to watch tonight, and I know the options, I can almost always predict what she will choose. Does this mean she isn't free? Not at all—she looks at what's playing, thinks over what she wants to see, and decides accordingly. No one is threatening her. No one is manipulating her or tricking her. No one has planted a remote-control device

in her brain. Thus, she chooses “of her own free will.” The fact that I can predict her choice changes nothing. Indeed, something would be wrong if I could *not* predict that she will prefer *The Blind Side* to *No Country for Old Men*.

The basic idea of Compatibilism may be summed up like this: Free will is compatible with causal determinism because “free” does not mean “uncaused.” Rather, it means something like “uncoerced.” You are free when you act according to your own wishes and desires.

Free Will as Involving Determinism. The main argument for Compatibilism goes as follows. The whole worry over free will begins with the idea that *if an action is part of the great causal chain, it cannot be free*. In other words, if human actions are like other events, subject to prior causes and controlled by the laws of physics, then we are no more free than a feather tossed about by the wind. The question, then, is whether this idea is true.

If no *caused* action could be free, then what could free actions be like? Presumably, they would have to be uncaused—they would have to be outside the great causal chain. But consider what that would mean. What would it be like for *any* event to be uncaused? Imagine that billiard balls stopped obeying the laws of physics. Their motions would then be unpredictable, but only because they would be random and chaotic. The balls might go off at odd angles, leap into the air, or suddenly stop. When struck by the cue ball, the eight ball might not move at all. Or it might explode, or turn to ice. Anything could happen.

Similarly, if a person’s actions were suddenly disconnected from the network of causes and effects, they would become random, chaotic, and unpredictable. A man standing on a street corner might step into traffic rather than wait for the light to change. He might attack the dog next to him as he recites the Magna Carta. This is what it might be like for behavior to be uncaused. But it is not what we mean by “free.” You would not think that someone who behaved in this way had free will—you would think that he was crazy. Free actions are not random and chaotic.

According to Compatibilism, freedom is not only compatible with Determinism; freedom *requires* Determinism, at least in the realm of human action. In a random, chaotic world, no one would be free, because free actions must be orderly and thoughtful. But in a world that operates according to the Laws

of Nature, free action is possible. In such a world, a person's character and desires can control what he does.

The Problem with Compatibilism. Compatibilism is a successful theory, in that it makes sense of how people actually talk: we do indeed say that people act “of their own free will” when they act according to their own wishes and values. However, historically, the debate about free will has been about something deeper than how people talk: it has been about the nature of human action and its relationship to moral responsibility. Peter van Inwagen (1942–) makes the deep case against Compatibilism like this:

If Determinism is true, then our actions are consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.

The compatibilist might respond by saying that, even if Determinism is true, our actions could still be “up to us” in the sense that we could perform those actions according to our own desires and values. However, van Inwagen's argument can be extended to our desires and values themselves: If Determinism is true, then our desires and values are themselves consequences of the laws of nature and events in the remote past. But those laws and those events are not up to us. Therefore, our desires and values are not up to us. This is a point that compatibilists must concede. But that concession seems to give away the game. If my actions depend on my desires and values, and my desires and values are not up to me, then my actions seem not to be up to me, either. So, free will does not appear to be compatible with Determinism after all.

9.4. Ethics and Free Will

Many philosophers and theologians see the deterministic implications of modern science as a crisis. Our freedom, they say, is essential to our dignity as moral beings. It separates us from the animals. If we start thinking of ourselves as mere robots, pushed around by impersonal forces, we lose our humanity.

But before we give in to such fears, we need to ask what the implications of Determinism really are. If we do not have free will, are we still responsible moral agents? Does ethics lose its point?

Robots, Fatalism, and Deliberation. We can set aside the idea that if we lack free will, then we are “mere robots.” We are nothing like robots. We have thoughts, intentions, and emotions. We experience happiness and unhappiness. We love our children and, if we are lucky, they love us back. We take pleasure in going to parties, playing Xbox, and listening to music. None of this is true of robots.

Also unlike robots, we often have *reasons* for what we do, and this will be true even if we lack free will. As long as our beliefs and desires can guide our behavior, we can act rationally. We can still pursue our own goals, just as before. Of course, the sense in which our goals are “our own” would change. We could no longer think of them as something that we freely choose. Instead, we would see them as the result of our genes, our environment, and the working of our brains. But so what? They would still be our goals, and we would still care about them.

It is sometimes suggested that the denial of free will would lead to a fatalistic attitude about the future: There would be no point in striving to change things, because the future must follow a set path. But this does not follow. The future depends on what we do, and if we want a certain sort of future, we have good reason to bring it about. Suppose you want sick children in Niger to get medical care, and so you contribute to humanitarian efforts. You help change the future. And there is certainly a point to it—without the help, the children will be worse off. The presence or absence of free will makes no difference.

Could we *deliberate* about what to do if we had no free will? Some philosophers think that, if we’re not free, then “deliberating” makes no sense. After all, deliberating means trying to decide, and trying to decide seems to presuppose that we could do different things. This reasoning sounds plausible. But what do we actually do when we deliberate? Mainly, we think about what we want and about how different actions would lead to different outcomes. We think about the children in Niger, what it’s like to be sick and helpless, how our money could supply their needs, and so on; and we might think about other things

we could use our money for. Nothing in all that presupposes freedom.

Nor does the denial of free will mean the end of ethics. We may still regard some things as good and others as bad—even if no one has free will, it is still better for the children in Niger not to die. We may still regard actions as better or worse depending on their consequences—contributing to humanitarian efforts is a good thing, even if we lack free will. And we can still think about all this in deciding what to do.

Evaluating People as Good or Bad. Can we continue to regard *people* as good or bad if they lack free will? In a sense, we certainly can. Even without free will, people will still have virtues and vices. They will still be brave or cowardly, kind or cruel, generous or greedy. A murderer will still be a killer, and a killer will still be a bad thing to be. Of course, it may be possible to explain someone's misdeeds as the result of his genes, his history, or the chemistry of his brain. And this may lead us to view him as unlucky. But that doesn't mean he is not bad. We need to distinguish *whether* someone is bad from *how he came to be* bad. A causal account of someone's character doesn't imply he isn't bad. It merely explains how he got to be that way.

Consider again Eric Rudolph, whose terrorist bombing spree in the 1990s killed two people and injured over 100 others. Rudolph's life story provides ample evidence that he is not responsible for how he turned out. As one journalist put it, Rudolph was "the product of a paranoid fringe of white supremacists, religious zealots and government haters." Knowing his background, we may regard him as unlucky. As the old saying has it, there but for the grace of God go I. Yet we may still think of Rudolph as bad, because he is, after all, a murderer. He deliberately tried to harm innocent people. These things are true even if we understand what made him that way.

Responsibility. But, it may be protested, if people do not have free will, then they are not *responsible* for what they do. Philosophers disagree about whether this conclusion would be disturbing or enlightened. Bertrand Russell (1872–1970) thought it enlightened. He wrote:

No man treats a motorcar as foolishly as he treats another human being. When the car will not go, he does not

attribute its annoying behavior to sin; he does not say, “You are a wicked motorcar, and I shall not give you any more petrol until you go.” He attempts to find out what is wrong and to set it right.

Similarly, Russell says, when a person misbehaves, we should try to figure out why and deal with that.

However, it is not clear whether the conclusion is correct—perhaps we can be responsible even if we’re not free. So let’s see if we can develop an account of responsibility that goes along with Determinism. Being responsible means, at least, that you may be held accountable for what you do—you may be blamed when you behave badly and praised when you behave well. So, if you are a responsible being, there must be some conditions under which you deserve blame for an action. There seem to be three such conditions: (a) you must have done the act in question, (b) the act must have been wrong, and (c) you must have no excuse for having done it.

The notion of an excuse is crucial. Excuses are facts that get you off the hook when you have done something bad. It was an accident, you may say, or you didn’t know what you were doing, or you were forced to do it. Here are some common excuses:

- *Mistake.* When you left my apartment, you took my umbrella by mistake—you thought it was yours. If you had taken my umbrella intentionally, you would deserve blame.
- *Accident.* You were driving safely, taking every precaution, when a child darted in front of the car, and you hit her. If you had been trying to hit her, or if you had been driving carelessly, you could be blamed.
- *Coercion.* You were forced to open the company safe because the robbers threatened you. If you had opened it voluntarily, you could be blamed.
- *Ignorance.* You gave your boyfriend a deadly poison because the bottle was mislabeled. If you had known it was poison, you’d be a murderer.
- *Illness.* You mistreated your sister because you suffer from Capgras syndrome. This is a rare delusional disorder that makes people believe that someone they know has been replaced by an imposter. So, your behavior is not your fault.

The logic of praise is similar to the logic of blame. You deserve praise for an action if (a) you did it, (b) it was a good thing to do, and (c) there are no conditions present that are analogous to excuses. It is curious that we have no name for these analogous conditions. We have a word, “excuses,” for the conditions that make blame inappropriate; but we have no word for the conditions that make praise inappropriate. Yet, clearly, similar ideas apply. If you do something good, but you do it merely by accident or from ignorance, then you do not deserve praise. Perhaps there is no general word for these conditions because people are usually happy to receive praise, regardless of whether they deserve it. Let’s call these conditions “praise-busters.”

According to this account of responsibility, people are responsible for what they do if there are no excusing conditions or praise-busters present. Then, if they behave well, they merit praise; and if they behave badly, they deserve blame. Nothing in this account conflicts with the assumption of Determinism.

Is Something Still Missing? If we lack free will, there is still a sense in which we can be good or bad. Our actions can still have good or bad consequences, and we can still talk about whether a person has any of the standard excuses for her actions. But something feels missing here. Don’t we want to evaluate people in more ways than this? Without free will, we can evaluate people only as we now evaluate dogs. Some dogs are gentle and sweet; others are mean and vicious. In this vein, we speak of “good dogs” and “bad dogs.” However, we normally think of human beings as capable of a higher goodness—a *moral* goodness. People, we believe, can freely create value. They can freely choose to do what’s right or to do what’s wrong. Dogs can’t.

Next, consider the notion of responsibility. If you’re training a puppy to be a good companion, you treat her as if she were responsible: you punish her for behaving badly, and you reward her for behaving well. Moreover, as part of the training, you acknowledge some excuses on her behalf (she can’t be expected to resist eating meat that has been left out too temptingly), and you reject other alleged excuses (she *can* resist the urge to beg for meat at the table). However, none of

us would say that dogs are “really” responsible for what they do; rather, we would say that sometimes we treat dogs *as if they were responsible*, based on the benefits of doing so, even though we know that dogs are not responsible. If human beings lack free will, shouldn’t we say the same about humans—shouldn’t we say that people are not really responsible for what they do, even though it is often useful to treat them as though they were? People could be “really responsible” only if they had free will.

Also, without free will, it would be odd to criticize anyone for anything he has done. After all, he had to do it. For example, suppose that a member of the president’s inner circle lies at a congressional hearing. If there is no free will, then the Laws of Nature compelled him to lie—he couldn’t have done otherwise. To say that he should have been truthful would be absurd; it would be like saying that he should have jumped over the moon—both actions were impossible for him at that time. To say that the official ought to have been honest would thus violate the principle, *ought implies can*. This is the principle that if you *ought* to do something, then it must be true that you *can* do it. If we lack free will, then all moral criticisms would violate this principle, because it would never be true that we “could” have done something that we didn’t do.

Those who reject free will might respond in one of two ways. On the one hand, they might embrace the idea that no one can ever be rightfully blamed for their actions. This position is defensible, though it goes against common sense. On the other hand, they might interpret the ought-implies-can principle in this way: “To say that you ought to do X implies that you ‘can’ do X *in the sense that doing X is within your general physical power*.” Not lying *is* within our general physical power, so it makes sense to say that we ought not to lie. However, jumping over the moon is not within our general physical power, so it makes no sense to say that we ought to jump over the moon. By embracing the ought-implies-can principle in this form, we ensure that we will make moral judgments only about actions that we could conceivably influence—namely, actions that are within our general physical power. For example, we could conceivably influence someone not to lie (by affecting his brain chemistry in the right way), but we could never influence

someone to jump over the moon (no change in brain chemistry could make someone's legs that powerful). If we accept these assumptions, then we can still praise and blame people for their actions, and our use of the word "ought" will have a practical point, namely, to try to influence people's behavior. However, the whole business of praise and blame would sound more like manipulation—like a sophisticated form of dog training—than like the recognition of a deep moral truth about the other person's behavior.

A belief in free will, therefore, does seem necessary to retain our commonsense picture of moral agency. However, that picture might be incorrect.

Our Knowledge of the World around Us

As the poet said, “Only God can make a tree”—probably because it’s so hard to figure out how to get the bark on.

—WOODY ALLEN,
ON SEEING A TREE IN SUMMER (2007)

10.1. Vats and Demons

A human brain has inputs and outputs. The inputs are signals from the person’s eyes and ears and other sense organs. The brain receives those inputs and processes the information. The outputs are signals that the brain sends to various parts of the body. Those signals may cause us to breath, blink, laugh, walk up a flight of stairs, and so on.

Now imagine that we have better technology. Suppose that, in our advanced world, a scientist removes someone’s brain and keeps it alive in a vat of nutrients. The scientist hooks up this brain to a computer that supplies it with the kind of inputs that normally come from the eyes and ears. If the computer does its job well enough, then the person won’t be able to tell the difference. His brain will process the inputs in the same way it once processed inputs from the eyes and ears. His brain will then produce outputs, which are transmitted back to the computer. The computer will return new inputs, and the saga will continue.

In this scenario, the person whose brain is in the vat will have normal experiences. From his point of view, nothing special is happening. He will meet his friends, go to work, have dinner, and watch TV. Or, at least, he will believe he is doing those things. In fact, his life will be a computer-generated illusion.

Now suppose it is suggested that you are that person—you are a brain in a vat, and your “life” is only an illusion. You may think you are reading a chapter in a book right now, but in fact the computer is only creating the illusion that you are. This suggestion seems absurd, but how could you prove it wrong? You can do nothing, it seems, to prove that your life is real. After all, every experience you have, including the experience of trying to prove you’re not a brain in a vat, could be supplied by the computer. You might protest that the whole story is technologically impossible—brains cannot be kept alive in vats, nor can computers sustain such illusions. But perhaps that is true only in your artificial, deluded world. In the real world, outside your vat, such computers do exist.

The Matrix (1999), a movie starring Keanu Reeves, explores this possibility. In the movie, people’s brains are not removed, but their bodies are connected to a giant computer that achieves the same effect. The people in the Matrix believe that they live in a physical world of buildings and weather and cars, but in fact that world exists only in their minds. *The Thirteenth Floor* (1999) and *Total Recall* (1990) exploit similar themes.

Long before there were movies and computers, René Descartes (1596–1650) dreamed up a similar scenario. Rather than imagining brains in vats, Descartes imagined that a powerful “evil spirit” was intent upon deceiving him—a spirit with god-like powers that wanted to fool him about everything. This spirit can not only manipulate Descartes’ experiences, but it can also manipulate his beliefs. If such an Evil Demon tried to deceive *you*, what would happen? Presumably, you would acquire false beliefs about everything. You would think that $3 + 3 = 5$, and that belief would seem as certain to you as $3 + 2 = 5$ now does. You might be sure about everything, yet right about nothing. And how could you possibly figure out what was really going on? Every thought you might have could be manipulated by the Demon.

Descartes also wondered how we can know we’re not dreaming. Often, we have vivid dreams in which we believe absurd things. Perhaps, right now, you are dreaming, and the thoughts you’re having are absurd, but you’re too confused to realize this. According to Descartes, “there are no conclusive indications by which waking life can be distinguished from sleep.” *Inception* (2010), a Christopher Nolan film, utilizes this theme. In that movie, characters are constantly trying to

figure out whether they're in a dream (or in a dream within a dream!), and doing so is not always easy.

The problem is not merely to say how we can *know for sure* that there's no Demon, that we're not dreaming, and that we're not brains in vats; the problem is how we can offer any evidence at all against these possibilities. Any alleged evidence we produce might just as easily be produced by a brain in a vat; and any confidence we have on these matters might have been planted in us by the Evil Demon, or might be due to our confusion inside a dream.

We'll return to these ideas later. But first, let's consider another outrageous possibility, which some philosophers believe describes the real world.

10.2. Idealism

When you look at a tree, what exactly do you see? One answer is that you see a patchwork of colors, mainly shades of brown and green. If you move your head, the pattern shifts slightly. Close your eyes, and the pattern disappears. The tree is still there, of course, but the thing you were experiencing—the splash of colors—will vanish. Open your eyes, and it will reappear.

The seventeenth-century philosophers called this pattern of colors an *idea* in your mind. In the twentieth century, the term *sense-datum* was popular. Today, philosophers speak of *qualia*. But whatever term is used, the point is that what you literally see is not the tree itself. What you literally see is a representation in your mind. This representation changes when you look at the tree from different angles. But the tree, presumably, remains the same. A color-blind person might have a different experience when looking at the tree. Rather than green and brown, she might see shades of purple.

But now a startling possibility arises. If we experience only ideas, then why do we need “physical objects” as part of our worldview? What do they add? We are never aware of them. Of course, we might say that physical objects are the *cause* of our ideas: we experience a patchwork of brown and green colors because there is a tree out there, causing us to have these experiences. But that hypothesis can never be proved. We can never step outside of our ideas to find their cause: no matter how we look at the tree (or touch it, or smell it), we will encounter only more ideas. The physical world is thus an unnecessary

supposition, so we might as well forget about it. *The only things that really exist are our minds and their ideas.*

This conclusion was embraced by George Berkeley (pronounced BARK-lee), an eccentric but brilliant Irish bishop who lived from 1685 to 1753. His writings have a vigor and incisiveness that have earned them a permanent place in intellectual history. The theory that reality consists entirely of minds and their ideas is called “Berkelian Idealism” in his honor—if it is an honor. Berkeley’s summation of this view is admirably clear:

though it were possible that solid, figured, movable substances may exist without the mind . . . yet how [could we] know this? Either we must know it by sense or by reason. As for our senses, by them we have the knowledge only of our sensations, ideas, or those things that are immediately perceived by sense, call them what you will; but they do not inform us that things exist without the mind, or unperceived. . . . It remains therefore that if we have any knowledge at all of external things, it must be by reason, inferring their existence from what is immediately perceived by sense. But what reason can induce us to believe the existence of [physical objects], from what we perceive, since the very patrons of matter themselves do not pretend there is any necessary connection [between physical objects] and our ideas?

Idealism strikes most of us as absurd because it denies that physical objects exist. But aside from insulting the view, what can we say against it? A number of objections come to mind. When I look at the tree outside my window, why do I always experience the same splash of colors? You and I can say, “Because you’re always looking at the same tree.” What can an idealist say? He doesn’t believe there is a tree. Or, when you and I both look out my window, why do we both see the same kind of thing? Again, we can say it’s because we’re seeing the same tree; but the idealist must call it a coincidence. And again: When I start the dishwasher and later come back to unload it, it appears as though something has been going on in my absence—the dishes have been getting cleaned. But what can Berkeley say? On his view there is no dishwasher, and there are no dishes. These objections to Idealism all boil down to the same thing: It seems that there is a world independent of us—a world that affects the same person similarly at different times, that affects

different people similarly, and that keeps going when we're not perceiving it.

Bishop Berkeley responded to these objections by bringing in God. God is responsible for ordering our ideas so that we can make sense of the world. When we are both looking through the same window, God makes sure we have the same experiences. And when I come back to unload the dishes, God makes sure that the dishes look clean. In fact, God saw them being cleaned—even when no one is looking at the tree or the dishes, they still exist as ideas in God's mind. God is always perceiving everything.

If, in response, the “materialist” challenges Berkeley to prove that God exists, Berkeley might offer some of the traditional arguments for God's existence. Or he might say, “To make sense of my theory, I have proposed that we accept the existence of one thing in addition to minds and ideas—namely, God. However, believers in an independent, physical reality are proposing that we accept the existence of countless additional things—namely, every single physical object in the universe. My hypothesis is simpler.”

10.3. What Evidence for These Views Might Be Like

So far, we've been discussing ideas that almost no one believes are true. There are no idealists, despite Berkeley's arguments. And, aside from a few lunatics, nobody believes they're a brain in a vat or that an Evil Demon controls their thoughts. However, views like these could turn out to be true.

Suppose that, when you finish reading this paragraph, there is a sudden discontinuity in your experience. In an eye blink, you find that you are in a hospital bed. Across the room is a mirror, and in the mirror you see an unfamiliar person with bandages on his (or her) head. You are feeling confused and panicky, when a doctor comes in and explains what is going on. She says that your brain was just removed from a vat, where it had been connected to a computer for many years, and reinstalled in your head. You are given a tour of the facility, where you see other brains connected by wires to the colossal computer. Then, leaving the hospital, you find that you are in a city that you have never heard of and that your “hometown” does not exist on any

map. If you had those experiences, it would be reasonable for you to conclude that you had been a brain in a vat.

Or instead, suppose that, after you finish reading this paragraph, the world begins to tremble and everything around you dissolves—all the buildings, trees, and even the ground melts away, leaving nothing but a white-tinged background. Other than being confused, you feel no discomfort. Then it is all replaced by a different, strange-looking environment, which in turn disappears, to be replaced by still another environment. After this happens a few times, you hear a booming voice, coming from everywhere at once. The voice identifies itself as God, and it explains that He has been giving you a demonstration of His powers. There are no “physical objects,” He explains; they are unnecessary—why should He create physical objects to cause you to have experiences when He can give you the experiences directly? If all this happened, it would be reasonable for you to believe in Idealism. And if the booming voice had instead described a battle in which He had triumphed over the Evil Demon, then, given enough detail, it could be reasonable for you to accept the existence of Descartes’ evil deceiver.

If any of these things happened, you would not be forced to change your worldview. Instead, you could conclude that you had gone crazy. But it would at least be clear that you possessed evidence for Idealism, or for the Evil Demon, or for the brain-in-a-vat hypothesis. Of course, as things stand, we have no such evidence. However, we have not yet offered any reason to *reject* these crazy ideas, either. Can we come up with any?

10.4. Descartes’ Theological Response

Descartes was the first person to consider this question. He lived in the early seventeenth century, when modern science was very young. Descartes was excited about the development of science, and he wondered whether scientific knowledge had secure foundations. In fact, Descartes wondered if *any* of our beliefs were on firm footing, since science is just the rigorous application of common sense to the study of the world.

For our knowledge to be secure, Descartes thought, it must ultimately rest on foundations that cannot be doubted.

Thus, Descartes hoped to find an absolutely certain ground for human knowledge. And he thought he did, in the following chain of reasoning:

- (1) What can I know for sure? If I seem to see a fireplace in front of me, or even if I believe that $2 + 3 = 5$, I can't be sure of these things because an evil spirit might be controlling my thoughts. But there is one thing that I do know with absolute certainty: *I know that I am now having certain thoughts and experiences*. Even if it is nothing but an illusion, I know that I am having the *experience* "fireplace in front of me." I cannot be wrong about that. Similarly, even if an evil spirit is deceiving me, I still know that I am having the *thought* " $2 + 3 = 5$." I know my own thoughts and experiences, and you know your own thoughts and experiences. That's the one thing we can't be mistaken about.
- (2) If it is certain that we have thoughts and experiences, then it is certain that we exist. After all, if we did not exist, we could not be having those thoughts. Descartes expresses this inference as "I think, therefore I am"—or, in the Latin version, *Cogito, ergo sum*—one of the most famous propositions in the history of thought.
- (3) Among our ideas is one that stands out, namely, the idea of God. Our other ideas, such as the idea of a fireplace, are ideas of things that might not exist in reality. But the idea of God is different, because it is the idea of a perfect being, and so it is the idea of something that must exist in reality. Why? Because not existing in reality is incompatible with being perfect. Therefore, God must exist. (This is the Ontological Argument, which we discussed in chapter 2.)
- (4) We have now established that we exist, along with our thoughts and experiences, and that God, a perfect being, also exists. But God, if he is perfect, cannot be a deceiver. It follows that God could not have made us so that we would be systematically deceived about what the world is like. A perfect, truthful God would design our sense organs and intellect so that we would come to have true beliefs, not false beliefs.

- (5) It follows, then, that our senses and our powers of reasoning are reliable sources of knowledge about the world around us.

This argument leaves Descartes with a residual problem: If our faculties were designed and made by a perfectly good creator, why do we sometimes make mistakes? Error, he says, results from human action, not divine action. When we make mistakes, it is because we employ our faculties carelessly, or because other humans set out to deceive us.

Descartes' argument provides an answer to our earlier worries. If there is a perfect God, as it says in step (3), then there is no Evil Demon (or if there is, then the Demon's power to deceive us is kept in check by God). And if, as it says in step (5), our senses and powers of reasoning are reliable, then Idealism should be false, since our senses and powers of reasoning so naturally lead us to believe that physical objects exist. Finally, the dream hypothesis and the brain-in-a-vat hypothesis should not be true in Descartes' world, since a perfect being would protect us from a life of perpetual illusion.

Is Descartes' argument sound? Descartes' reasoning has been analyzed endlessly for the last three-and-a-half centuries. Today in North America there are probably a dozen graduate students writing dissertations about it. One problem is that serious objections have been raised to the Ontological Argument for God's existence. Another problem is that Descartes' procedure is circular: He reasons his way to the conclusion that reasoning can be trusted. Essentially, his argument goes like this: "How do we know that reasoning can be trusted? Because God created our powers of reasoning, and God is not a deceiver. How do we know that? Because we have a chain of reasoning to prove it." Among philosophers, this is known as "the Cartesian Circle." Because of these problems, nobody now thinks that Descartes provided an absolutely certain basis for human knowledge. That goal was too ambitious. Complete certainty, it seems, is beyond human powers.

Even if his argument was faulty, Descartes made a number of lasting contributions. First, he clearly identified the problem. Second, he showed that a theological perspective might provide a solution to that problem: human knowledge might find a secure basis in the existence of an honest deity. It is well known that

religion might inform our views on morality, mortality, and the origin of the universe. Surprisingly, Descartes shows that religion might also help explain the basis of knowledge. Indeed, we might say that Descartes presents us with a challenge: If we do *not* assume the existence of God, then how can we solve the problem?

10.5. Direct vs. Indirect Realism

We still haven't solved our central problem. We haven't explained why a belief in physical matter is more reasonable than a belief in Idealism, nor have we warded off Descartes' deceiving Demon or refuted the brain-in-a-vat scenario. Admittedly, our problem has an air of unreality about it. You might think it looks like a problem invented by people who have too much time on their hands. If so, many philosophers agree with you. On their view, the problem arises from a mistake.

So far, our discussion has assumed a particular view of how sense perception works. The view is this:

When we look at an object, we do not see the object itself. Instead, we are aware of an "experience," or an "idea in our minds," or a "sense-datum," or a "mental representation," or some such. Then we want to infer from these experiences that the object exists. The problem is to explain how this inference is justified.

But, critics say, this account of perception is mistaken. The mistake is saying that, when we look at a tree, we see only some sort of mental image. Instead, *we see the tree*. We see a physical object that exists outside us. We do not "infer" that the tree exists. We see it directly.

Seeing is a way of getting information about the world around us. But it is not a two-step process, in which we first get information about "sense-data" and then move from that to information about the tree. Instead, it is a one-step process of seeing the tree. That's how we know the tree is there. The brain-in-a-vat hypothesis, and the Evil Demon hypothesis, cannot be refuted, but under normal circumstances, we have good reason to reject them—we have direct awareness of the reality they deny.

Of course, we may sometimes be mistaken, but that fact poses no special problem. We discover our mistakes in the same way we learn everything else, by using our senses and our

intelligence. I think I see a tree—is it an illusion? I find out by looking more closely, or by trying to touch it, or by asking someone else to check, or by considering whether I have been drinking too much. If I think I might be suffering a hallucination but have no way to check, I may simply suspend judgment. Normally, however, I know I see a tree.

This view is known as Direct Realism. It is contrasted with Indirect Realism, which holds that our immediate object of perception is something mental—our idea of a tree—whereas we are only indirectly aware of the tree itself. Direct Realism appeals to common sense and seems to give a straightforward answer to the science-fiction scenarios. Its defenders include such eminent figures as Ludwig Wittgenstein (1889–1951) and J. L. Austin (1911–1960). But Direct Realism is unsatisfying, for two reasons.

First, Direct Realism doesn't actually help us solve problems like the brain in the vat. We want to be given evidence that we're not brains in vats; we want a reason to believe that we *are* in touch with an external reality. Direct Realism, however, merely asserts that we're in touch with an external reality. And, on the strength of that assertion, it concludes that we *know* we're in touch with an external reality. To borrow a phrase from Bertrand Russell, this solution to the problem has all the advantages of theft over honest toil.

Second, Direct Realism does not fit well with what we know about how the brain processes information. "Seeing" a tree is not a simple matter at all; it is the result of neurological processes that have a complicated structure. Our theory of perception should mesh with our scientific understanding of how this processing works.

10.6. Vision and the Brain

The German philosopher Immanuel Kant (1724–1804) is famous for insisting that perception is not a passive process. The mind does not simply record what passes before it; instead, the mind actively interprets experience according to certain built-in principles. Therefore, what we think of as "simple" perception is actually the result of a complicated interpretation of the sensory data. Today, nobody accepts the details of Kant's account. However, his basic idea is confirmed by current research.

The psychologist Steven Pinker (1954–) writes:

When [organisms] apprehend the world by sight, they have to use the splash of light reflected off its objects, projected as a two-dimensional kaleidoscope of throbbing, heaving streaks on each retina. The brain somehow analyzes the moving collages and arrives at an impressively accurate sense of the objects out there that gave rise to them.

How does the brain do this? Nobody knows for sure, but one leading idea is that the brain makes a number of assumptions about how the “throbbing, heaving streaks” are to be interpreted. Thus, what we see is determined by these assumptions as well as by the streaks. Here are four of those assumptions:

1. Sharp lines are interpreted as boundaries or edges, while enclosed areas of one color are taken to be surfaces.
2. Objects have the simplest shapes consistent with the pattern of light that hits our retinas. Thus, a skinny rectangle is more likely to be a stick than a penny viewed edgewise.
3. Surfaces are assumed to be evenly textured. Gradual changes in surface coloration are therefore interpreted as due to lighting and perspective. For example, if you see a surface covered with rows of dots, and if the dots in succeeding rows seem to be closer and closer together, your brain will interpret this as an effect of perspective. It will “see” the dots as evenly spaced. Likewise, if the surface grows gradually darker from one side to the other, the surface will be perceived as evenly colored, with the shading interpreted as an effect of the lighting.
4. To quote Pinker: “Objects have regular, compact silhouettes, so if Object A has a bite taken out that is filled by Object B, A is behind B; accidents don’t happen in which a bulge in B fits flush into the bite of A.” Think of a suitcase with a bowl in front of it, so that you can’t see the corner of the suitcase. You don’t assume that the corner of the suitcase is missing and that the bowl fits neatly into the vacant space. Instead, you assume that the bowl is in front of the suitcase, blocking your view.

These assumptions are what philosophers call *defeasible*—they are correctable in light of further experience. Thus, we might move the bowl and discover that the suitcase really does have a missing corner. Or, if the sticklike object is rotated, it could turn out to be a penny.

A further problem is why we don't perceive the world as a montage of two-dimensional surfaces. Instead, we perceive it as a collection of three-dimensional *objects*. How do we manage that? Psychologist Irving Biederman argues that the brain interprets information from the eyes using a stock of 24 basic three-dimensional shapes, which he calls "geons." Biederman suggests that the geons are the brain's inbuilt device for constructing objects from the data of experience. The geons include such basic units as a sphere, a cube, a cylinder, and a bent tube; and all objects are said to be constructed by combining these basic units. (This will come as no surprise to art students, who traditionally are taught to draw and shade such standard "solids" as a preliminary to tackling more complicated objects.) A bent tube on top of a box is a suitcase. A bent tube on top of a cylinder is a bucket. A bent tube on the side of a cylinder is a cup. And so on. Kant spoke of the mind imposing its forms on experience; according to Biederman, geons are among the forms.

Color, one of the last elements added by the brain, can make a huge difference to our emotions. The neurologist Oliver Sacks (1933–) has studied what happens to people who suffer brain damage that makes them completely color-blind, so that they see only whites, blacks, and greys. For some of these people, vision is not like watching a black-and-white TV—rather, something important is missing along with the colors. One of Sacks's patients, Jonathan I., was a painter who became color-blind after a car accident. His eyes weren't damaged, but the part of his brain that constructs color was. Even in his mind's eye, he could see only blacks, whites, and greys. According to Sacks, Jonathan I. found his new world to be alien, empty, and dead. He avoided other people, who seemed to him "like animated grey statues," and he found food to be disgusting. Even when he closed his eyes, Jonathan I.'s mental image of a tomato looked just as black as its appearance.

Finally, we have two eyes that view the world from slightly different angles, and the brain has a way of combining the information from them to tell us how far away objects are. To us,

it may seem obvious that perceiving depth is the purpose of binocular vision. However, this is a recent discovery. Until the nineteenth century, it was thought that having two eyes is simply a by-product of the fact that our bodies are symmetrical; or, it was thought, the second eye may be a spare in case something happens to the first.

It should thus come as no surprise that most of the fibers going to the brain's vision center do not come from the retina. Rather, 80 percent of those fibers come from deep within the brain itself. Our brains work hard to interpret the splashes of light reflected onto our retinas. Are those interpretations arbitrary? Why do we make them? The answer is that the human brain is the product of evolution. We process information in the ways that contributed to the survival of our ancestors, who needed to find food, avoid predators, interact with other members of their species, and otherwise move safely through the world. Thus, we have evolved a perceptual system that is useful for our purposes.

10.7. Conclusion

The commonsense view of perception goes something like this:

We have experiences such as “seeing a tree” or “hearing a cricket” because our bodies interact with a physical world that includes things like trees and crickets. The world impinges on our sense organs, causing us to have experiences and beliefs that represent the world to us in a fairly accurate way. The physical world exists independent of us—that is, it would exist even if we didn't exist, and it continues to exist even when we are not observing it.

Nothing we've discussed gives us any reason to reject this commonsense view. However, in the previous section, we saw that the brain plays a more active role in perception than common sense might suppose. The brain does not merely “read off” the data of experience coming from our senses; rather, the brain adds to those data and actively interprets them. As one writer put it, “The account of perception that's starting to emerge is what we might call the ‘brain's best guess’ theory of perception: perception is the brain's best guess about what is happening in the outside world.”

According to common sense, Idealism is false: we really do interact with a physical world. Also, common sense holds that our experiences and beliefs “represent the world to us in a fairly accurate way.” Thus, common sense rejects the idea that we’re being systematically deceived, either by Descartes’ Demon or by the scientists who manipulate all the brains in vats. However, these are mere assumptions—the commonsense view offers no argument against Idealism, nor does it give us any reason to believe that we’re not being systematically duped. What, then, has become of our attempt to refute these bizarre possibilities? It has failed. We have found no reason at all to believe that we’re not living in the Matrix.

Ethics and Objectivity

There are no objective values.

—J. L. MACKIE, *ETHICS: INVENTING RIGHT AND WRONG* (1977)

11.1. Thrasymachus's Challenge

Thrasymachus has the misfortune of being remembered through the eyes of someone who despised him. He was a Sophist, one of the professional teachers who flourished in Athens during the time of Socrates (ca. 470–399 B.C.). We know about Socrates through the writings of his student Plato. Plato tells us that the Sophists, unlike Socrates, charged a fee for their instruction, and he is quick to insinuate that the Sophists liked money more than truth. Plato is especially hard on Thrasymachus, who is introduced in the *Republic* like this:

While we had been talking [says Socrates] Thrasymachus had often tried to interrupt, but had been prevented by those sitting near him, who wanted to hear the argument concluded; but when we paused, he was no longer able to contain himself and gathered himself together and sprang on us like a wild beast, as if he wanted to tear us in pieces. Polemarchus and I were scared stiff, as Thrasymachus burst out and said, “What is all this nonsense, Socrates?”

The “nonsense” was a discussion of justice. Thrasymachus was impatient because Socrates and his friends were assuming that justice is something real and important. But according to Thrasymachus, people believe in right and wrong only because they are taught to obey the rules of their society. Those rules, however, are merely human inventions. Thrasymachus added that the rules of a society will protect the interests of the society's

most powerful members. So, when ordinary people think they must “do the right thing,” they are just being played.

Throughout history, there have always been groups of people who, like Thrasyarchus, believe that ethics is just a matter of opinion; and there have always been groups of people who, like Socrates, believe that ethics has an objective basis. But with the rise of modern science, skepticism about ethics became even more attractive. Modern science sees the world as a cold, indifferent place that cares nothing for us or our projects; the universe is a realm of facts that know nothing of right or wrong. As David Hume (1711–1776) put it, “The life of a man is of no greater importance to the universe than that of an oyster.” Thus, it seems natural to conclude that ethics can be nothing but a human creation.

Thrasyarchus challenged Socrates to prove that ethics has an objective basis. One way of meeting Thrasyarchus’s challenge might be to introduce religious notions. If the universe was created by God, according to a divine plan, and if God issues commands about how we should live, we might find in this an objective basis for our judgments of right and wrong. But suppose we set this possibility aside. Is there any way to defend the objectivity of ethics without invoking religion? We will see that there is a way. The arguments for ethical skepticism are not as powerful as they appear.

11.2. Is Ethics Just a Matter of Social Conventions?

The idea that ethics is nothing more than social convention has always appealed to educated people. Different cultures have different moral codes, it is said, and it is merely naïve to think that there is one universal standard that applies in all places and times. Examples of differences are easy to come by. In Islamic countries, men may have more than one wife. In medieval Europe, lending money for interest was considered a sin. In northern Greenland, old people were sometimes left to die in the snow. Considering such examples, anthropologists have long agreed with Herodotus, the ancient historian who said, “Custom is king o’er all.”

Today, the idea of morality as a social product is attractive for an additional reason. Given America’s economic power and military might, it is especially important for Americans to respect

and appreciate the differences between cultures. In particular, it is said, we must avoid the arrogant assumption that our ways are “right” and that other ways are “wrong.” This means, in part, that we should refrain from making moral judgments about other cultures. We should adopt a policy of live and let live.

On the surface, this attitude seems enlightened. Tolerance is important, and many cultural practices are nothing more than social customs—standards of dress, the details of household arrangements, the methods of greeting, and so on. But fundamental matters of justice are different. When we consider such examples as slavery, racism, and the abuse of women, it no longer seems enlightened to shrug and say, “They have their customs, and we have ours.” Consider two examples.

In a Pakistani village, a 12-year-old boy was accused of being romantically involved with a 22-year-old woman of a higher social class. He denied it, but the tribal elders did not believe him. As punishment, they decreed that the boy’s teenage sister—who had done nothing wrong—should be publicly raped. Her name is Mukhtar Mai. Four men carried out the sentence while the village watched. Observers said there was nothing unusual in this, but with so many foreigners in the region, the incident was noticed and reported in *Newsweek*. This was in 2002.

In the same year, in northern Nigeria, a religious court sentenced an unwed mother named Amina Lawal to be stoned to death for having had sex out of wedlock. When the verdict was read, the crowd in the courtroom cheered, and the judge said that the sentence should be carried out as soon as Lawal’s baby was old enough not to need breast-feeding. Lawal identified the father, but he denied the accusation, and no charges were brought against him. This was only one in a series of such sentences imposed in northern Nigeria. Responding to international pressure, the Nigerian government announced that it would not enforce the sentence against Lawal, and in 2004 she was set free.

The rape of Mukhtar Mai was regarded as a matter of tribal honor. Her brother was allegedly romancing a woman from a different tribe, and the elders of her tribe demanded justice. The stonings in Nigeria, on the other hand, are the application of the Islamic law of Sharia, which has been adopted by 12 of Nigeria’s 36 states. Our instincts are to condemn both punishments. But are we *justified* in condemning them? Two thoughts stand in the way of saying we are.

First, there is the idea that *we should respect the differences between cultures*. People in other cultures have a right to follow their own traditions, it is said, even if outsiders like us disapprove. After all, their traditions may have a purpose that we don't understand. Moreover, it is said, their values might be different from ours. Should thoughts like these persuade us to hold our tongues when we hear about a raping or a stoning on the other side of the world?

They should not. Respecting a culture does not mean tolerating everything in it. You might think that a culture has a wonderful history and has produced great art. You might think that its leading figures are admirable and that your own culture has much to learn from them. Yet, despite all this, you need not regard the culture as perfect. You might think it contains elements that are awful. Most of us take just this attitude toward our own society—if you are an American, you probably think that America is a great country but that some aspects of our culture are appalling. Why should you not think the same about Pakistan or Nigeria? If you did, you would be agreeing with many Pakistanis and Nigerians.

Moreover, it is a mistake to think of the world as a collection of discrete, unified cultures that exist in isolation from one another. Cultures overlap and interact. In the United States, there are cultural differences between Irish Catholics, Italian Americans, Southern Baptists, African Americans in Los Angeles, African Americans in Mississippi, and Hasidic Jews in Brooklyn. Coal miners in West Virginia are quite different, culturally, from stockbrokers in New York City. In some ways we think that “live and let live” is the best policy, but no one takes this to mean that you should have no opinion about what happens in another part of the country.

This also raises the question of who speaks for a culture. Is it the priests? The politicians? The women? The slaves? Opinions within a society are rarely uniform. If we say, for example, that slavery was approved of in ancient Greece, we are referring to the opinions of the slave-owners. The slaves themselves might have had a different opinion. Or consider again the public raping of Mukhtar Mai. When this happened, the Pakistani government took action against the tribal leaders who had ordered it. Which group—the local leaders or the national government—sets the standards that we must respect? There are no clear-cut answers to these questions. Thus, it is often unhelpful to say, “We must respect the values of the culture.”

Finally, we should notice a purely logical point. Some people think that ethical relativism *follows from* the fact that cultures have different standards. That is, they endorse this argument:

- (1) Different cultures have different moral codes.
- (2) Therefore, there is no such thing as objective right and wrong. Where ethics is concerned, the standards of the different societies are all that exist.

But this is a mistake. It does not follow from the fact that people disagree about something, that there is no truth about it. For example, cultures may disagree about the Milky Way—some think it's a galaxy, others think it's a river in the sky—but it does not follow from this that there is no objective truth about the Milky Way. The same goes for ethics. When cultures employ different customs, some of the customs may be better than others. It is easy to overlook this if we only think of trivial examples, such as the standards of dress at a wedding or funeral. Those may indeed be nothing but matters of local custom. But it does not follow that *all* practices are merely matters of local custom. Fundamental matters of justice might be different.

Thus, we needn't refrain from morally assessing the customs of other societies. We can be tolerant and respectful, yet think that other cultures are flawed.

There is, however, a second reason why being judgmental may seem inappropriate: all standards may seem to be culture-relative. If we say that the rape of Mukhtar Mai was wrong, we are using *our* standards to judge *their* practices. From our point of view, the rape was wrong, but why is our point of view correct? We can say that the tribal leaders are wrong, but they can equally well say that we are wrong. How can we get beyond this mutual finger-pointing?

This second argument can be spelled out as follows:

- (1) For our criticisms of other cultures to be justified, they must appeal to standards that are not simply derived from our own culture.
- (2) But there are no such culture-neutral moral standards. All standards are relative to some society or other.
- (3) Therefore, our criticisms of other cultures are unjustified.

Is this correct? It sounds plausible, but in fact there *is* a culture-neutral standard of right and wrong. The reason we object to the rape and the stoning is not that they are “contrary to American values.” Nor is our objection that these practices are somehow bad for *us*. We object because of the harm done to Mukhtar Mai and Amina Lawal. Our culture-neutral standard is *whether the social practice in question is beneficial or harmful to the people who are affected by it*. Good social practices benefit people; bad social practices harm people.

This standard is culture-neutral in every relevant sense. First, it does not play favorites between cultures. It may be applied equally to all societies, including our own. Second, the source of the principle does not lie within one particular culture. Rather, every culture values the welfare of its people. It is a value that must be embraced, at least to some extent, by every culture, if the culture is to exist. And so, the suggestion that a social practice is harmful can never be dismissed as an alien standard “brought in from the outside.”

11.3. Ethics and Science

We think of science as the most objective human enterprise. Scientists know how to get at the truth. The study of ethics, however, seems unlike science. So how can ethics be objective? Let’s consider three arguments along these lines.

The Argument from Disagreement. It is troubling that ethical disagreement seems so widespread and persistent. If ethics were a matter of objective truth, shouldn’t we expect more consensus? Yet it seems that, in matters of ethics, people disagree about everything. They argue over abortion, capital punishment, gun control, euthanasia, the environment, and the moral status of animals. They disagree about sex, drug use, and whether we have a duty to help needy children in foreign countries. In science, however, there is widespread agreement on all essential points. The natural conclusion is that ethics, unlike science, is a mere matter of opinion. We may summarize the argument like this:

- (1) In ethics, unlike in science, there is widespread and persistent disagreement.

- (2) The best explanation of this fact is that there is no objective truth in ethics.
- (3) Therefore, we may conclude, at least tentatively, that there is no objective truth in ethics.

Is this argument sound? We may begin by observing that ethics is more like science than people think. There is a tremendous amount of agreement in ethics. All thoughtful people agree that murder, rape, theft, child abuse, blackmail, kidnapping, and racism are wrong. Everyone agrees that we should tell the truth, help our friends, keep our promises, and love our children. If it is said that some people do not agree—racists and thieves, for example—it can be replied that some people disagree with the findings of science—flat-earthers and psychics, for example. The situation in ethics is the same as in science: The vast majority of people agree, while some dissenters are ignored, for good reason. In fact, there may be more dissenters in science than in ethics, if you count religious fundamentalists who reject Darwin's theory of evolution.

In ethics, then, there is massive agreement about simple matters. But there is also disagreement about abortion, capital punishment, and the other issues mentioned above. What should we make of this? We might notice, first, that from a social point of view, most of the matters we disagree on are less important than the matters we agree on. No matter how much people care about gun control, for example, gun control is less important than murder, truth telling, or promise keeping. After all, societies can function with a variety of policies on guns. But social living would be impossible without a rule against murder. Likewise, society would be impossible if people could lie and break their promises at will. To see this, try to imagine what it would be like to live in a place where people could lie, cheat, steal, and kill with impunity. That society would collapse.

We might also notice that many of the disputed issues are *harder* than the matters we agree on. In order to take a principled stand on abortion, for example, we would have to figure out whether the fetus's life is precious. Thus, we would have to answer the question: When, in the course of human development, does human life acquire its full value? Moreover, we might have to assess the importance of potentiality: Does the fact that the fetus *might one day* become a full human person affect the fetus's moral

status *today*? All this is hard enough, but to make things worse, it is not obvious that these are even the right questions to ask. We could instead ask about the rights of the pregnant woman, or about how large the human population should be. So, it is no wonder that people disagree about the morality of abortion. In such cases, the difficulty of the issues, and not the absence of “truth,” is the best explanation of why people don’t agree.

A similar pattern of agreement and disagreement exists in science. All scientists agree about a large central core of accepted truth. Yet many others issues are disputed. Scientists disagree about the path that evolution has taken, the prospects for string theory, what is really shown by the groundbreaking experiments on infant cognition, and the relation between quantum theory and classical relativity. So, contrary to superficial impressions, there does not appear to be any fundamental difference between ethics and science. Both are characterized by broad agreement alongside some disagreement.

The Argument from Lack of Proof. While scientists may disagree about some things, they agree about how to resolve their disputes. They make observations and perform experiments in order to prove or disprove their hypotheses. This means that, in science, disagreement is only temporary. But in ethics, disagreement is endless, because no one knows how to prove or disprove anything.

This argument may be summarized as follows:

- (1) If there were objective truth in ethics, then it should be possible to prove that at least some ethical opinions are true.
- (2) But it is not possible to prove that any ethical opinion is true.
- (3) Therefore, there is no objective truth in ethics.

Is this correct? It certainly sounds plausible. Anyone who has tried to persuade someone else on an ethical matter will know how frustrating it can be. A pacifist, for example, will not be persuaded that violence is sometimes necessary, no matter what reasons are offered. Nor will a defender of violence be persuaded to change his mind, no matter how often he is told that violence only begets more violence.

If we turn to simpler examples, however, things look very different. Suppose the issue is whether a certain doctor is unethical. I say that Dr. Jones behaves shamefully, and you are surprised to hear it because you think she's a fine doctor. So I point out several things:

- Dr. Jones owns stock in a drug company, and she always prescribes that company's drugs, whether her patients need them or not.
- She won't listen to the advice of other physicians, and she becomes angry when her patients want a second opinion.
- She doesn't keep up with current medical knowledge.
- She often performs surgery while she's drunk.

Suppose all this is true. Isn't this good evidence that she is unethical? Doesn't this *prove* that she is unethical? Suppose, further, that little could be said on the other side, in her defense. Doesn't this settle the matter? What more in the way of proof could anyone want?

Other examples come easily to mind. The proof that Mr. Smith is a bad man is that he is a habitual liar who is sometimes cruel. The proof that Mr. Brown is an unethical poker player is that he cheats. The proof that Professor Adams should not have given the midterm on Tuesday is that she announced it would be given on Thursday. In each case, of course, further facts may need to be considered. But the point is that such judgments are not merely "subjective."

Ethical proofs may be different, in some ways, from scientific proofs. But that does not mean that ethical proofs are deficient. Ethical proofs consist in giving reasons to support moral conclusions. If the reasons are powerful, and if there are no good opposing considerations, then the case is made.

This may seem too quick. If ethical judgments can so obviously be proven, then why was the contrary idea so plausible in the first place? Why is it so intuitively appealing to think that there are no ethical proofs? There are at least three reasons.

First, when we think about ethics, we don't usually think about the simple matters. The fact that they are simple makes them boring, so we tend to ignore them. We are attracted instead to the harder issues, such as pacifism, abortion, and gun control. But this is what leads us astray. If we think only about the hard issues, we may naturally conclude that there are no

proofs in ethics, because no one has a knockdown argument against abortion or in favor of pacifism. We might overlook the fact that proofs are easily available on the more mundane issues.

Second, there are often good reasons on both sides of a moral issue, and this leads people to despair about reaching definite conclusions. If I say that Smith is a bad man because he often lies, you may reply that Smith sometimes works for good causes. The first fact counts against him, but the second is in his favor. This feature of moral reasoning, however, should not worry us. Moral reasoning requires taking all the facts into account and weighing them against one another. Where Smith is concerned, the right conclusion might be that he is on the whole bad, even though he has some good qualities. Or the right conclusion might be that Smith is equally good and bad. It just depends on the facts. The difficult issues are like this; there is much to be said on each side.

That simple truth is often overlooked. It is common for people on one side of a debate to deny that the other side has made any good points at all. To win the debate, they feel that they must concede nothing. This, of course, means that the debate will go on endlessly. But moral debates needn't be like that. You can acknowledge that your opponent has made good, relevant points, while maintaining that your points are even better.

Finally, it is easy to confuse *proving an opinion to be true* and *persuading someone to accept your proof*. The first is a matter of sound reasoning; the second is a matter of psychology. Someone might reject a perfectly good argument because he is stubborn, or prejudiced, or simply uninterested in finding the truth. It does not follow that the argument itself is defective. A Klansman or a neo-Nazi may not listen to a sound argument about racism, but that says something about *them*, not about the argument. And there is a more general reason why people might resist listening. Accepting a moral argument often means acknowledging that we should change our behavior. And we may not want to do that. Thus, people sometimes turn a deaf ear.

The Metaphysical Argument. There is one further argument to consider, namely, that ethics cannot be objective because “values” do not exist as part of the objective world.

If we take an inventory of the world, noting all the things that exist, we can make a very long list, mentioning rocks, rivers,

mountains, plants, and animals. We would find buildings, deserts, caves, iron, and oxygen. Looking up, we would see stars, comets, clouds, and galaxies. Of course, we could never complete such a list. Life is too short, there are too many things, and we are too ignorant. But we think we know, at least roughly, the *kinds* of things that exist. There are physical objects, made of atoms, that obey the laws of physics, chemistry, and biology; and there are conscious beings, such as ourselves, which may or may not be just another type of physical object.

But where, among all these things, are values? The answer, it seems, is *nowhere*. Values do not exist, at least not in the same way that rocks and rivers exist. Apart from human feelings and human interests, the world appears to contain no values. David Hume put it this way:

Take any action allow'd to be vicious: Willful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In whichever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of [disapproval], which arises in you, towards this action. Here is a matter of fact; but 'tis the object of feeling, not of reason.

Of course, there may be other conscious beings who also have feelings and interests—nonhuman animals, for example, and possibly the inhabitants of other planets. But they will be in the same position that we are in. They will find no values in the world around them. Only their “passions, motives, volitions, and thoughts” will give rise to values for them.

Other philosophers have taken up this theme. Friedrich Nietzsche (1844–1900) was a troubled figure who had no real home for much of his life. He moved around Europe, writing book after book, but his books were largely ignored. Eleven years before his death, he went insane. After he died, his fame grew, but he was embraced by the Nazis, who wrongly took him to support their racist causes. Later he was rescued from this misunderstanding, and today Nietzsche is seen as an original and important thinker.

Like Hume, Nietzsche denied that there are moral facts. “There are no moral phenomena at all,” he wrote, “but only

moral interpretations of phenomena.” The right way to think about ethics, he thought, is not to focus on *morality*, as though it were one unified thing, but to study *moralities*, the historically contingent systems of value that have been created by different people at different times. Nietzsche himself wrote about a morality that underlies much of Western culture, which he called “slave morality.” Slave morality, Nietzsche said, glorifies meekness, self-denial, obedience, and poverty. This outlook, he said, is unworthy of noble men, and he advocated replacing it with an ethic that emphasizes strength and dominance.

This is the “Metaphysical Argument”: *Ethical opinions cannot be objectively true or false because there is no moral reality that they may match or fail to match.* This is the deep way in which ethics differs from science. Science describes a reality that exists independent of observers. If conscious beings ceased to exist, the world would otherwise be unchanged—it would still be there, and it would still be just as science describes it. But if there were no conscious beings, there would be no moral dimension to reality at all.

We may summarize the argument like this:

- (1) There are objective truths in science because there is an objective reality—the physical world—which science describes.
- (2) But there is no moral reality comparable to the reality of the physical world. There is nothing “there” for ethics to describe.
- (3) Therefore, there are no objective truths in ethics.

Is this argument sound? It is true, I think, that there is no moral reality comparable to the reality of the physical world. However, it does not follow from this that ethics has no objective basis. After all, an inquiry might be objective in two ways:

1. It may be objective because there is an independent reality that it describes correctly or incorrectly. Science is objective in this sense.
2. An inquiry may be objective because there are reliable methods of reasoning that determine truth and falsity in its domain. Mathematics is objective in this sense. Mathematical results are objective because they are provable by the relevant kinds of arguments.

Ethics is objective in the second sense. We do not discover whether an ethical opinion is true by comparing it to some sort of “moral reality.” Instead, we discover what is right by examining the reasons, or arguments, that can be given on both sides of an issue—the right thing to do is whatever there are the best reasons for doing. For ethics to be objective, it is enough that we can identify and evaluate reasons for and against ethical judgments.

11.4. The Importance of Human Interests

The preceding discussion should have dispelled most of your doubts about the objectivity of ethics. But some nagging doubts might remain, for good reason. We have not yet gotten to the bottom of things.

Every inquiry, whether in science, mathematics, or ethics, involves reasoning: We gather data, marshal arguments, and draw conclusions. But reasoning cannot go on forever. If I tell you that A is true, and you want to know why, I may cite B as my reason. If you call B into question, I may justify B by appealing to C. And so on. But at some point we come to the end of the line. Every chain of reasoning must end somewhere. This means that every argument ultimately appeals to some consideration that is simply taken for granted.

Scientific reasoning terminates when we reach simple facts about the physical world. We know, for example, that the galaxies are moving apart. How do we know this? Because of facts about the light that reaches the earth—in particular, facts about the red shift in the spectrum. How do we know what the red shift means? Because of many past observations and experiments. This example is simplified, but when we reach the simplest observed facts, we have reached the bedrock on which everything else rests. Mathematical reasoning is somewhat different in that it does not appeal to facts about the physical world. Instead, it relies on axioms, which may be taken as self-evidently true or may simply be assumed for the purposes of the proof.

Where do ethical arguments terminate? To what do they ultimately appeal? Let us look more closely at one of our previous examples. Smith is a bad man because, among other things, he is a habitual liar. This is a good reason for judging him to be morally deficient, we said, and so this fact forms part of a “proof” that he behaves unethically. But why is it bad to lie?

Lying is bad for several reasons. First, it is harmful to people. If I lie to you, and you believe me, then things can go badly for you. Suppose you ask me when the concert starts, and I say “Ten,” even though I know it begins at 7:30. You arrive at 9:40 only to find that you’ve missed half the show. Multiply this example many times over, and you will see why honesty is important. Second, lying is a violation of trust. When you believe me without checking on what I say, you are trusting me. So, if I lie to you, I am causing you harm in a special way, by taking advantage of your trust. This is why people feel so personally affronted when they’re lied to. Finally, the rule against lying is a fundamental social rule, in the sense that no society could exist without it. If we cannot assume that people will speak truthfully, then communication cannot take place; and without communication, society cannot exist.

Thus, the judgment that lying is wrong is not arbitrary. It has good reasons behind it. Suppose, though, you pushed further and wanted to know why it matters if people are harmed, or trust is violated, or society collapses. We could say a little more. We could point out that people are *worse off* when they base their decisions on false information and their trust is violated. We could point out that everyone would be worse off if people couldn’t live together in cooperative societies. But you persist: Why does it matter if people are worse off? Here we come to the end of the line. Ethical reasoning terminates in considerations about people being better or worse off—or, perhaps, in considerations about *any* sensitive creature being better or worse off—just as scientific reasoning terminates in simple observations about the physical world.

Some people conclude from this that ethics is subjective—after all, ethics is ultimately about the well-being of individuals, which is a subjective matter. Personally, I don’t think this is a good use of the word “subjective,” but what’s important is not what words we use, but what conclusions we draw from them. If ethics is subjective in the sense of being about individuals, it does not follow that ethics is arbitrary. Nor does it follow that people are free to accept whatever ethical judgments they like, or that one person’s opinions will always be as good as another’s. Ethics remains a matter of following reason, and it will still be objectively true that some things are good for people, and other things are bad for people. Ethical judgments can still be correct or incorrect. In these ways, the objectivity of ethics is secure.

Why Should We Be Moral?

Immorality is the morality of those who are having a better time.

—H. L. MENCKEN,
A MENCKEN CHRESTOMATHY (1956)

12.1. The Ring of Gyges

An ancient legend tells the story of Gyges, a poor shepherd who found a magic ring in a fissure opened by an earthquake. Gyges discovered that when he twisted the ring on his finger, he would become invisible. With this great power, Gyges could go anywhere and do anything he wanted, without fear of detection. He used the ring to enrich himself, taking what he wanted and killing anyone who got in his way. Eventually, he invaded the royal palace, where he seduced the queen, murdered the king, and seized the throne.

Glaucon tells this story in Book I of Plato's *Republic*. The story is meant to illustrate how behaving immorally can sometimes be to one's advantage. If Gyges had remained virtuous, he would have remained poor. By breaking the moral rules, however, he became rich and powerful. Considering this, why should Gyges care about the moral rules? For that matter, why should any of us be moral if it doesn't serve our own needs? Why tell the truth, if lying is more convenient? Why give money to charity, when you can spend it on yourself? Morality places restrictions on us that we may not like. So why shouldn't we just forget about it? Glaucon adds that, in his opinion, all of us would behave like Gyges, if we thought we could get away with it.

In what follows, we will take up the question of why we should be moral. But first, we need to understand the question. It is not a request for a justification of moral behavior. If it were,

then the answer would be easy. We could easily say why Gyges shouldn't have robbed and murdered his way to the throne. Robbery is taking things that do not belong to you, and murder inflicts a terrible harm on its victims. Similarly, it is easy to explain why we should be truthful or why we should give money to help the needy. Lying harms people, and hungry people need food more than people like us need money.

Such reasons, however, only determine what is right, and that is not the issue. Glaucon's challenge arises after the moral reasoning has been done. We may grant that it is morally right to respect people's lives and property. We may concede that it is right to tell the truth and help people. Glaucon's question is, *Why should we care about doing what is right?*

To answer this question, we must show that living morally is in our own best interests, and that will not be easy. On the surface, it looks like ethics is a hindrance in promoting one's own happiness. Of course, it may be a good thing for you if *other people* live ethically, for then they will respect your rights and be helpful to you. But if *you* are bound by moral constraints, that is another matter.

Can it be shown that accepting moral constraints really is in one's self-interest? We might be able to meet Glaucon's challenge, and we might not. It all depends on what morality is.

12.2. Ethics and Religion

One familiar idea is that right living consists in obedience to God's commands. On this conception, God has set out the rules we must obey, and he will reward those who follow his rules and punish those who do not.

If this were true, we could answer Glaucon's challenge as follows: We should be moral because otherwise God will punish us. Even if, like Gyges, we had the power of invisibility, we would still be visible to God, and so, ultimately, we could not get away with being wicked. In one familiar scenario, the righteous will spend eternity in heaven while the sinners will go to hell. Thus, any benefit you might gain from wrongdoing will be only temporary. In the long run, virtue pays.

That response, however, may not fully answer Glaucon's challenge. On the familiar scenario, God does not reward individuals in proportion to their virtue. Rather, God sorts

everyone into just two groups: the heaven-bound and the hell-bound. In the end, both the saints and those who barely scrape by wind up in the same heavenly paradise. Thus, it might be to your advantage to do just enough to get into heaven, but no more. And so, you can ask: Why should I be moral, when a little immorality will benefit me on earth without spoiling my chances at heaven?

On these assumptions, Glaucon's challenge can be partly answered: Even if we can get away with some immorality, we shouldn't risk hellfire by robbing and murdering our way to the throne. But in fact, much more virtue than that might be called for. In examining our own lives, we can never be too sure as to what God might be thinking. You never know when one more selfish act might tip the scales. Thus, we should probably be as good as we can be, to increase our chances of eternal bliss.

Let's now discuss, more generally, the moral theory that puts God in the center of things. According to the Divine Command Theory, an action is morally right if, and only if, it is commanded by God. Is this theory correct? If we tried to live by it, we would encounter enormous practical difficulties. How can we know what God commands? Some people claim that God has told them what he wants us to do. But why should we trust those people? Hearing voices can be a sign of schizophrenia, and anyway, they might be lying. They might be saying they've spoken to God merely to get attention, or to feel important, or to persuade their followers to send them money.

Others rely, more modestly, on Scripture and Church tradition for guidance. But those sources are notoriously difficult to interpret. They give vague and sometimes contradictory instructions. So, when people consult these authorities, they typically rely on their own judgment to sort out what seems acceptable. For example, they may cite the passage in Leviticus that condemns homosexuality while ignoring the passage that requires you to wash your clothes if you touch anything a menstruating woman has sat upon.

But these are just practical difficulties. It may still seem plausible that God's commands provide the ultimate basis for ethics: God's saying that something is wrong is what *makes it* wrong. Many religious people think it would be sacrilege not to accept this view. Socrates, however, gave a powerful argument against it.

In the *Euthyphro*, Socrates considers whether “right” can be the same as “what the gods command.” Socrates accepted that the gods exist and that they may issue instructions. But he showed that this cannot be the ultimate basis of ethics. As he said, we need to distinguish two possibilities: Either the gods have good reasons for their instructions, or they do not. If they do not, then their commands would be arbitrary—the gods would be like petty tyrants who demand that we do this and that even though there is no good reason for it. This is an impious view that religious people will not want to accept. On the other hand, if the gods do have good reasons for their instructions, then there must be a standard of rightness independent of their commands—namely, the standard which they themselves adhere to in deciding what to require of us.

It follows, then, that the rightness or wrongness of actions cannot be understood merely in terms of their conformity to divine commands. We may always ask why the gods command what they do, and the answer to that question will reveal why right actions are right and why wrong actions are wrong.

The same is true of sacred texts. Nothing can be morally right or wrong simply because a book says so. If the book’s decrees are not arbitrary, then there must be some reason for them. For example, the Bible says we should not lie about our neighbors—we should not “bear false witness” against them. Is this rule arbitrary? Certainly not. Lying causes harm and violates the trust that others have in us, and lying about our neighbors is insulting to them. *That’s* why lying is wrong. The reason is not, “because the Bible says so.” Similarly, we may ask why homosexuality is condemned. Are there good reasons for this pronouncement? If so, then those reasons will give the real explanation of why homosexuality is wrong. If not, then homosexuality isn’t wrong, and the Bible’s condemnation is unjustified.

These problems make the Divine Command Theory implausible, but they do not refute the separate idea that, if God punishes wrongdoing, then we have good reason to be moral. This idea so impressed the great German philosopher Immanuel Kant (1724–1804) that he made it into an argument for God’s existence. Kant reasoned that if God does not exist, then the universe is morally incomplete, because virtue will go unrewarded and wickedness will go unpunished. This thought was

intolerable to him, so he concluded that God exists. Even great philosophers, it seems, can indulge in wishful thinking.

12.3. The Social Contract

In the seventeenth century, with the rise of modern science, philosophy became an increasingly secular enterprise. Since then, there has been a rough consensus that ethics must be understood as a human phenomenon—as the product of human needs, interests, and desires.

Thomas Hobbes (1588–1679) was the first great modern thinker to offer a secular basis for ethics. Hobbes assumed that “good” and “bad” are what we call the things that we like or dislike. Thus, when you and I like different things, we regard different things as good or bad. However, Hobbes said, we are all essentially alike. We are all self-interested creatures who want to live as well as possible. This is the key to understanding ethics. Ethics arises when people realize what they must do to live as well as they can.

Hobbes points out that each of us is enormously better off living in a cooperative society than trying to make it on our own. The benefits of social living go far beyond companionship. Social living makes possible schools, hospitals, and highways; houses with electricity and central heating; airplanes and cell phones; books, magazines, and websites; movies, opera, and football; and science, engineering, and agriculture. Without social cooperation, we would lose all of this. Therefore, it benefits each of us to establish and maintain a cooperative society.

But it turns out that a cooperative society can exist only if we adopt certain rules of behavior—rules that require telling the truth, keeping our promises, respecting one another’s lives and property, and so on:

- If people couldn’t be relied on to tell the truth, then no one would care what anyone said. Communication would be impossible. And without communication, society would collapse.
- If people didn’t keep their promises, then there could be no division of labor—workers could not count on getting paid, retailers could not rely on their suppliers, and so on—and the economy would collapse. There could be no business, no building, no agriculture, and no medicine.

- Without assurances against assault, murder, and theft, no one could feel secure. Everyone would have to be constantly on guard, and no one could trust their neighbors.

Thus, to obtain the benefits of social living, we must strike a bargain with one another, whereby each of us agrees to obey these rules. This “social contract” is the basis of morality. Thus, morality may be understood as *the rules that a self-interested person would agree to obey, provided that others agree to obey them as well.*

Why the Social Contract Theory Is Attractive. This way of thinking about morality has a number of appealing features. First, it takes the mystery out of ethics and makes it a practical, down-to-earth business. Living morally is not a matter of blind obedience to the mysterious dictates of a supernatural being. Instead, it is about doing what it takes to make social living possible.

Second, the social contract approach gives us a sensible way of determining what our moral duties are. We are morally required to do the things that make social living possible. Otherwise, we may do what we like. Unfortunately, when many people hear the word “morals,” they think about an attempt to restrict their sex lives. But an ethic based on the social contract would have little interest in people’s personal affairs.

Third, the Social Contract Theory gives a plausible answer to Glaucon’s question: Why should we behave morally when it is not to our advantage? Well, it *is* to our advantage to live in a society in which people behave morally. Thus, it is rational for us to accept moral restrictions on our conduct, because the overall system benefits us. There is, however, a problem with this answer.

The Problem of the Free Rider. Schemes of social cooperation always face the “free rider” problem. A free rider is someone who benefits from a cooperative arrangement without contributing to it. Suppose the home owners in my neighborhood chip in to pay for a streetlight. I want the light as much as anyone else does, but I refuse to chip in, knowing that they will go ahead without me. The light is installed, and I get to use it for free. That makes me a free rider.

In any cooperative society, there will be some free riders—individuals who benefit from living in the society, but who don’t

follow the rules which make the society possible. Glaucon's challenge may therefore be reformulated: Why not be a free rider? Each of us already enjoys the benefits of living in a stable society. So, why shouldn't we break the rules if we think we can get away with it?

The free-rider problem for ethics can be solved, but only partially. The partial solution goes like this: Each of us has good reason not only to encourage others to obey the social rules but also to make it as hard as possible for them to break those rules. Take the rule against murder, for example. You don't merely want to encourage others not to murder you. You want a situation in which *no one can get away with* murdering you. Each of us has good reason to support the creation and maintenance of a social system in which other people could not murder us. To accomplish this, we establish laws and other methods of enforcement. But in doing so, we create a situation in which *we* cannot get away with murder, either.

I said that this is a "partial" solution, for two reasons. First, we use the power of the law to enforce the rules against murder, theft, and other grave offenses, but not all social rules are suitable for legal enforcement. Rules of ordinary decency must be enforced in "the court of public opinion" rather than in the court of law. The penalty for lying, for example, is only that people will get angry at you. Although everyone wants to avoid prison, not everyone minds upsetting the people around them. Thus, informal methods of enforcement are bound to be only partly effective.

Second, no mechanisms, formal or informal, are going to be perfect. It is easy to get away with the occasional lie. One might even get away with the occasional murder. Glaucon's challenge remains: Why should you obey the moral rules when you think you can get away with breaking them?

12.4. Morality and Benevolence

The Social Contract Theory does not assume that people are altruistic. Each person can be motivated to obey the social rules out of simple self-interest. However, people are not entirely selfish. Human beings have at least some benevolent feelings, if only for family and friends. We have evolved as social creatures just as surely as we have evolved as creatures with lungs. Caring for our kin and members of our local group is as natural for us as breathing.

If humans do have some degree of natural altruism, does this have any significance for ethics? David Hume (1711–1776), the great Scottish philosopher, thought so. Hume agreed with Hobbes that our moral opinions are expressions of our feelings, but he did not believe that our feelings are merely self-centered. He believed that we also have “social sentiments”—feelings that connect us with other people and make us care about them. That is why, Hume says, we measure right and wrong by “the true interests of mankind”:

In all determinations of morality, this circumstance of public utility is ever principally in view; and wherever disputes arise, either in philosophy or common life, concerning the bounds of duty, the question cannot . . . be decided with greater certainty than by ascertaining . . . the true interests of mankind.

This view came to be known as Utilitarianism. In modern moral philosophy, it is the chief alternative to the Social Contract Theory.

Utilitarianism. Utilitarians believe that one principle sums up all our moral duties: *We should always try to produce the greatest possible benefit for everyone who will be affected by our action.*

This Principle of Utility is deceptively simple. It actually combines three ideas. First, in determining what to do, we should be guided by the consequences of our actions—we should do whatever can be reasonably expected to have the best outcome. Second, in determining which consequences are best, we should care only about the benefits and harms that would be caused—we should do whatever will cause the greatest benefits and the least significant harms. And third, the Principle of Utility assumes that each person’s happiness is as important as anyone else’s.

Although Hume suggested this idea, two philosophers from England pursued it in greater detail. Jeremy Bentham (1748–1832) led a movement to reform the laws of Britain along utilitarian lines. The Benthamites were remarkably successful in advancing such causes as prison reform and restrictions on the use of child labor. John Stuart Mill (1806–1873), the son of a Benthamite, gave the theory its most popular and influential defense in his book *Utilitarianism*.

The utilitarian movement attracted critics from the outset. It was an easy target because it ignores conventional religious notions. Morality, according to the Utilitarians, has nothing to do with obedience to God or gaining credit in heaven. Rather, the point is just to make life in this world as comfortable and as happy as possible. Thus, some critics condemned Utilitarianism as a godless doctrine. To this Mill replied:

[T]he question depends upon what idea we have formed of the moral character of the Deity. If it be a true belief that God desires, above all things, the happiness of his creatures, and that this was his purpose in their creation, utility is not only not a godless doctrine, but more profoundly religious than any other.

Utilitarianism was also an easy target because it rejected many conventional moral notions. Bentham argued, for example, that the purpose of the criminal justice system cannot be understood in the traditional way as “paying back” offenders for their wicked deeds—that only piles misery upon misery. Instead, the social response to crime should be threefold: to identify and deal with the causes of criminal behavior; where possible, to reform individual lawbreakers and make them into productive citizens; and to “punish” people only insofar as it is necessary to deter others from committing similar crimes. Today these ideas are familiar, but only because the utilitarian movement was so successful. Or, to take a different example: By insisting that everyone’s happiness is equally important, the Utilitarians offended various elitist notions of group superiority. According to the utilitarian standard, neither race nor sex nor social class makes a difference to one’s moral status. Mill himself wrote a book called *The Subjection of Women* that became a classic of the nineteenth-century suffragist movement.

Finally, Utilitarianism was controversial because it has no use for “absolute” moral rules. According to Utilitarianism, the traditional moral rules are merely “rules of thumb”—they’re good pieces of advice, but they admit of exceptions. Whenever breaking a rule will have better results than following the rule, we should break it. The rule against killing, for example, might be suspended if someone is dying of a painful illness and requests a painless death. Moreover, the Utilitarians regard some traditional rules as suspect, even as rules of thumb. Christian

moralists had traditionally said that masturbation is evil, but from a utilitarian point of view, masturbation is a good thing. A more serious matter is the religious condemnation of homosexuality, which has resulted in misery for countless people. Utilitarianism implies that if an activity makes people happy, without anyone being harmed, it cannot be wrong.

Utilitarianism says that our moral duty is to promote the general happiness. Why should we do that? Mill echoes Glaucon's challenge when he says, "I feel that I am bound not to rob or murder, betray or deceive; but why am I bound to promote the general happiness? If my own happiness lies in something else, why may I not give that the preference?" Aside from the "external sanctions" of law and public opinion, Mill saw only one possible reason for accepting this or any other moral standard. The "internal sanction" of morality, he said, must always be "a feeling in our minds." And the kind of ethic we accept, he thought, will depend on the nature of our feelings. If human beings have "social feelings," then Utilitarianism will be the natural standard for them:

The firm foundation [of utilitarian morality] is that of the social feelings of mankind—the desire to be in unity with our fellow creatures, which is already a powerful principle in human nature, and happily one of those which tend to become stronger, even without express inculcation, from the influences of advancing civilization.

Impartiality. Utilitarianism has implications that conflict with traditional morality. Much the same could be said about the Social Contract Theory. In most of the practical matters that we have mentioned—criminal punishment, racial discrimination, women's rights, euthanasia, homosexuality—the two theories agree. But they differ dramatically on one issue. Utilitarians, but not social contract theorists, think that we have an extensive moral duty to help others.

Suppose, for example, you are thinking of spending \$1,000 for a new living room carpet. Should you do this? What are the alternatives? One alternative is to give the money to an agency such as the United Nations Children's Fund (UNICEF). Around 8.8 million children under the age of 5 die each year, mostly from preventable causes such as pneumonia, diarrhea, malaria, and birth asphyxia. By giving your money to UNICEF

and making do with your old carpet, you could provide life-saving medical care for dozens of children. From the point of view of utility—seeking the best overall outcome for everyone—there is no doubt that you should give the money to UNICEF. Obviously, the medicine will help the children far more than the new rug will help you.

The Social Contract Theory takes a different approach. If morality rests on an agreement between people—an agreement we enter into to promote our own interests—what would the agreement say about helping other people? It depends on whether those other people can help us. If they can, then we could benefit from an agreement to help each other. If not, then we have no reason to accept any restrictions on our conduct.

From this point of view, we would have no reason to accept a general duty to provide aid to children in foreign countries. Jan Narveson, a social contract theorist, says that we needn't "go very far out of our way to be very helpful to those we don't know and may not particularly care for." And, Narveson asks,

What about parting with the means for making your sweet little daughter's birthday party a memorable one, in order to keep a dozen strangers alive on the other side of the world? Is this something you are morally required to do? Indeed not. She may well *matter* to you more than they. This illustrates again the fact that people do *not* "count equally" for most of us. Normal people care more about some people than others, and build their very lives around those carings.

Which view is correct? Do we have a moral duty to help strangers or not? Here is a thought experiment that might help. Suppose there are two buttons on my desk, and I must choose which to press. By pressing button A, I can provide my daughter with a memorable party; by pressing button B, I can save the lives of a dozen strangers. Is it really all right for me to press A just because I care more for my daughter? What would your "conscientious feelings" tell you? Mill believed that one's conscientious feelings—the feelings that prevail after everything has been thought through—determine one's obligations. He believed that we cannot, when we are thoughtful and reflective, approve of pushing button A.

However, some contemporary Utilitarians have argued that the matter need not be left to the uncertainties of

individual feeling. Their argument goes like this: It may be true that we all care more for ourselves, our family, and our friends than we care for strangers. But we have rational capacities as well as feelings, and if we use those, we will realize that there are no relevant differences between us, those close to us, and strangers. Strangers have needs and interests, just like we do. Thus, no one should take his or her own well-being to be especially important. Peter Singer (1946–), a utilitarian philosopher, writes:

Reason makes it possible for us to see ourselves in this way. . . . I am able to see that I am just one being among others, with interests and desires like others. I have a personal perspective on the world, from which my interests are at the front and centre of the stage, the interests of my family and friends are close behind, and the interests of strangers are pushed to the back and sides. But reason enables me to see that others have similarly subjective perspectives, and that from “the point of view of the universe,” my perspective is no more privileged than theirs. Thus my ability to reason shows me the possibility of detaching myself from my own perspective, and shows me what the universe might look like if I had no personal perspective.

So, from an objective viewpoint, each of us must acknowledge that our own perspective—our own particular set of needs, interests, likes, and dislikes—is only one among many and has no special status. Morally, everyone counts equally, even strangers in foreign countries who have no ability to help us or harm us.

Conclusion. We have reached no firm conclusions about what morality requires of us. However, it looks like morality requires us to care about the needs and interests of total strangers. But why should we do that? Glaucon’s challenge remains: Why should we do what’s right, if there’s nothing in it for us?

A religious outlook that includes a belief in the afterlife could help us answer Glaucon’s question. Otherwise, we can appeal to the fact that everyone benefits from a social arrangement in which the moral rules are acknowledged and enforced. But the free-rider problem cannot be completely solved; people can always get away with a certain amount of bad behavior. In the end, we can only hope that people’s behavior will be guided by what Mill called “the conscientious feelings of mankind.”

This may seem like a feeble conclusion. Yet “the conscientious feelings of mankind” are a powerful force, made stronger by education and the advancement of civilization. Still, we must concede that, if people can get away with wrongdoing and genuinely don’t care about others, then nothing will stop them. We could remind them of all the reasons why their actions would be wrong and remind them that their interests don’t matter more than the interests of other people. But that will only prove that their actions *are* wrong. To stop them from behaving badly, something more is required, namely, that they care about those reasons.

The Meaning of Life

Death's at the bottom of everything.

—GRAHAM GREENE, *THE THIRD MAN* (1950)

13.1. The Problem of the Point of View

In 1826 John Stuart Mill became obsessed with the thought that his life was meaningless. He was only 22 years old, and his great works were still ahead of him, but he had already begun to make a name for himself. He wrote for a radical journal founded by Jeremy Bentham, the great utilitarian philosopher, and he often spoke on behalf of progressive causes. “I had what might truly be called an object in life,” Mill said, “to be a reformer of the world.” But then he suddenly lost confidence. His work no longer seemed important to him. He came to believe that, even if he achieved everything he wanted, he would not be happy. “I was in a dull state of nerves,” Mill said, “... I seemed to have nothing left to live for.” This condition persisted for a year while outwardly he carried on as if nothing were wrong. But inwardly, the clouds grew darker.

Finally, as often happens in such cases, things turned around. Reading about a boy's reaction to his father's death, Mill was moved to tears, and he found his love of life returning:

Relieved from my ever present sense of irremediable wretchedness, I gradually found that the ordinary incidents of life could again give me some pleasure; that I could again find enjoyment, not intense, but sufficient for cheerfulness, in sunshine and sky, in books, in conversation, in public affairs; and that there was, once more, excitement, though of a moderate kind, in exerting myself for my opinions, and for the public good. Thus the cloud gradually drew off, and I again enjoyed life.

One might understand Mill's experience as merely pathological: He was depressed, but eventually he snapped out of it. And that is probably true. But the meaning of life is a real problem that can arise even when we're not depressed. It is actually more than one problem, because different issues might be at stake. For Mill, the question was how to be happy. A different question is whether there is anything truly worth living for.

At the deepest level, the problem of the meaning of life arises from a clash between two points of view, each of which is natural and unavoidable for us. On the one hand, each of us occupies a personal point of view from which our lives and projects seem immensely important. We care about our family, our work, and whether the Red Sox will win the pennant. We have desires, goals, and plans. We are distressed by the fact that we will die, and on our deathbeds we regard that prospect as a calamity.

On the other hand, we can step outside our individual points of view and look at things from an impersonal standpoint. From "the point of view of the universe," our lives have little significance. What does it matter if I raise a family, succeed in my work, or take a trip to London? Eventually, I will die, and everything I do will come to nothing. In fact, the whole human race will vanish one day, leaving no trace behind.

The contrast could not be greater. From your own point of view, who you are and what you do are surpassingly important. But to the universe you are nothing. What are we to make of this? As Thomas Nagel (1937–) puts it, "In seeing ourselves from outside we find it difficult to take our lives seriously. This loss of conviction, and the attempt to regain it, is the problem of the meaning of life."

13.2. Happiness

The ancient philosophers had a lot to say about happiness. They assumed that human beings strive to be happy and that happiness consists in a life of reason and virtue. Epicurus (341–270 B.C.) recommended plain living, so as to avoid pains and anxieties. The Stoics added that a man should not let his happiness depend on things he can't control, like wealth, health, good looks, and the opinion of others. We cannot control external events, they said, so we should be indifferent to them and take life as it comes. Epictetus (ca. A.D. 55–135), one of the great Stoic teachers, told his students, "Ask not that events should

happen as you will, but let your will be that events should happen as they do.”

In recent years, happiness has been a hot topic in social science. In a typical “happiness study,” the researcher will conduct a large survey, asking people how happy they are as well as their age, sex, race, religion, income level, educational history, political affiliation, and so on. Then the researcher will use statistical tools to look for significant results among the data. Using this method, investigators have discovered, for example, that the elderly tend to be happier than people in their 20s and 30s, and men tend to be happier than women.

Researchers have also discovered that happiness is largely genetic: Genes account for a whopping 80 percent of the variation in human happiness. That finding is confirmed by studies which show how *unimportant* environmental factors can be. Winning the lottery, for example, seems to have no long-term effect on happiness. People are ecstatic for a while, but in a few days the euphoria subsides: Grumpy people return to being grumpy, and cheerful people come down from the clouds but remain happy. The same is true of disasters. It took only three weeks for car crash victims in Michigan to become happy again, despite having suffered permanent, debilitating spinal cord injuries. Cancer patients also bounce back: They are unhappy upon learning their diagnosis, but soon they return to whatever level of happiness is normal for them.

Does wealth affect happiness? Everyone agrees that being impoverished is bad: People who lack necessities are less happy, on average, than those who have them. But suppose you have enough money to get by. Would having even more money make you happier? The answer seems to be, “Probably, but not as much as you might think.”

Money, in general, is correlated with happiness in three ways: richer countries have happier citizens than poorer countries; as countries become richer over time, their citizens become happier; and, at any given moment, the richer people in a society tend to be happier than the poorer people. Together, these facts strongly suggest that income affects well-being. However, the impact is not very great. Remember, first, that genes account for about 80 percent of the variation in human happiness. Therefore, all the other factors, including wealth, can only account for about 20 percent. Second, recall that winning the lottery doesn’t seem to provide any long-term boost in happiness.

If money made a huge difference to well-being, then presumably we could discern a difference in people's happiness before and after winning the lottery. Finally, these generalizations have exceptions. For example, Americans are four times richer than Mexicans, but Mexicans report being happier. Americans are three times richer than they were 50 years ago, but according to the surveys, they are no happier.

Happiness is also correlated with other things. One is personal control: People are happier being their own boss, for example, rather than taking orders from someone else. Another factor is good relations with others, especially friends and family. A third factor is meaningful work: People who feel that they are accomplishing something worthwhile report being especially happy. In general, happiness is self-sustaining, because happy people tend to behave in ways that keep them happy. They have higher opinions of other people and are more outgoing. Thus, they form friendships more easily. Happy people are also more optimistic. They tend to choose long-term rewards over short-term satisfactions.

You might think, then, that in order to be happier you need to take control of your life, make friends, and seek out meaningful work. All that may be true, but there is a catch. If you value these things only as a means to your happiness, then they will not make you happy. Happiness cannot be sought directly. Instead, you must value friends and work in their own right. Happiness is then the welcome by-product. As John Stuart Mill said, "Ask yourself if you are happy, and you cease to be so. The only chance is to treat, not happiness, but some end external to it, as the purpose of life."

13.3. Death

Some people believe that even a happy life can be meaningless. After all, they say, we will inevitably die, and anyway, the universe doesn't care about us. We'll consider these ideas in turn.

What attitude should we take toward our mortality? It depends, of course, on what we believe happens when we die. Some people believe they will live forever in paradise. Death, therefore, is like moving to a much better neighborhood. If you believe this, you should think that death is a wonderful thing, for you will be much better off after you die. Apparently, Socrates had this attitude, but most people do not.

On the other hand, death may be the permanent end of your existence. When you die, you're gone. It is important to understand what this means. Some people assume that nonexistence is a hard-to-imagine condition. They ask, "What is it like to be dead?" and they expect a profound answer. But that is a mistake. We cannot imagine what it is like to be dead because there is nothing to imagine. What your life will be like in 2210 is what it was like in 1810.

If death is the end, what attitude should we take toward it? Most of us find it awful. We hate the thought of dying, and we are willing to do almost anything to prolong our lives. Epicurus, however, said that we should not fear death. In a letter to one of his followers, he argued that "Death is nothing to us," because when we are dead we do not exist, and if we do not exist nothing can harm us. We will not be unhappy; we will suffer no pain; we will not be afraid, worried, or bored; and we will have no regrets. Therefore, Epicurus concluded, the wise person will not fear death.

There is something to this. However, it overlooks the fact that death is bad because it is a huge deprivation—if our lives could continue, we could enjoy all sorts of things. Thus, death is an evil because it puts an end to the good things of life. After I die, human history will continue, but I won't get to be a part of it. I will see no more movies, read no more books, make no more friends, and take no more trips. If my wife survives me, I will not get to be with her. I will not know my grandchildren's children. New inventions will appear and new discoveries will be made about the nature of the universe, but I won't know what they are. New music will be composed, but I won't hear it. Perhaps we will make contact with intelligent beings from other planets, but I won't know about it. That is why I don't want to die, and Epicurus's argument is beside the point.

But does the fact that I am going to die make my life meaningless? After all, it is said, what is the point of working, making friends, and raising a family if, in the end, we will all turn to dust? This thought has a certain emotional resonance, but it involves a fundamental mistake. We must distinguish the value of a thing from how long that thing will last. Something can be good while it lasts, even though it will not last forever. For example, in Afghanistan the Taliban destroyed a number of ancient monuments of great beauty and religious significance. Yet we would never say that these monuments lacked value, just because their time on earth was limited. A human life can also be wonderful, even though it must end.

At the very least, the mere fact that it will end does not extinguish whatever value it has.

13.4. Religion and the Indifferent Universe

From the point of view of the universe, it is said, human life has no significance. W. B. Yeats (1865–1939), the great Irish poet, put this matter in chilling perspective:

Where are now the warring kings,
 Word be-mockers?—By the Rood,
 Where are now the warring kings?
 An idle word is now their glory,
 By the stammering schoolboy said,
 Reading some entangled story:
 The kings of the old time are dead;
 The wandering earth herself may be
 Only a sudden flaming word,
 In clanging space a moment heard,
 Troubling the endless reverie.

Warring kings must die, the wandering earth is an insignificant speck, and the whole history of the human race is only a blip in time. In light of this, how can we attach any meaning to what we do?

One way to overcome this problem is to adopt a religious point of view according to which the universe is *not* indifferent to us. The great monotheistic religions—including Judaism, Christianity, and Islam—believe that God created the universe to provide a home for us and that our destiny is to live with Him forever. Human life is not, therefore, a meaningless blip in the history of the universe. Instead, we are the leading stars in the whole cosmic drama.

I will leave aside the question of whether the religious story is true, in order to ask a different question. Suppose it is true. How does the religious story help us understand the meaning of life? This question turns out to be puzzling.

Perhaps our lives are meaningful because *God has a plan for us*. But it is hard to see how the intentions of an outside agent could make your life meaningful to you. Suppose your parents have a plan for you: They want you to become a doctor. You might find this agreeable; but you might find it smothering, because you want to make your own plan. Indeed, if an outside

force dictates the terms of your life, you might regard your life as having no meaning at all.

Another possibility is that our lives are meaningful because *we are the objects of God's love*. According to the great religions, the universe isn't indifferent to us; rather, its Creator and Overseer cares about us deeply. But again, though this might be comforting, it is hard to see how it helps. If love is needed to give our lives meaning, most of us already have that. We have family and friends who care about us, and we care about them. If we think that our lives lack meaning despite this, how is the addition of someone else's love supposed to help?

Finally, from a religious point of view, human life is seen as a *permanent* feature of the universe. Death is overcome. But we have already considered this thought. The duration of a life is separate from its meaning. Indeed, an eternal life might be meaningless if it is devoid of everything that makes life worthwhile. And a short life might be full of worthwhile things.

Religion *can* help us understand the meaning of life, but so far we haven't considered a key element of religious belief. That element is the believer's own commitment. She willingly accepts her role as God's child and chooses to adopt a way of life that goes with it. This is important, because when she embraces her religion, it is no longer being imposed on her from the outside. Instead, those values become her own. She lives a religious life "from the inside," as a matter of conviction. Her life then gets its meaning from the things that she values: God's plan for her, God's love, and her own involvement in the religious life of her community.

But note that her life has meaning, not because she pursues *religious* values, but because she pursues *her own values*. And she might value something else. For example, she might see herself as a scholar, an athlete, a chess player, a musician, or a businessperson. Like Mill, she might strive "to be a reformer of the world." Then her life would get its meaning from the values associated with those activities, just as her life now gets its meaning from her religion.

Religion does have one great advantage over these other ways of finding meaning: It solves the problem of the indifferent universe. If, like Mill, you take it as your goal to reform the world, you must still face the fact that the universe doesn't care. That is why it is so tempting to say that, without God, life is meaningless. At the same time, however, the religious commitment has a drawback that the other commitments do not

have: It assumes that the religious story is true. We do not want to base our lives on moonshine; if the religious story is not true, we are doing just that.

13.5. The Meaning of Particular Lives

There is a difference between *life's* having a meaning and *a life's* having a meaning. If the religious story is true, then human life in general has a meaning. Without religion, we can only say that each life has its own particular meaning and that what our lives mean is up to us.

The comedian Groucho Marx once said, "Those are my principles. If you don't like them, I have others." Hopefully, your own principles mean more to you than Groucho's meant to him. Embracing your values will always add meaning to your life, if you are truly embracing them. *What* you value will depend on your talents, dispositions, and beliefs. Think of Socrates, St. Francis, Yogi Berra, Jonas Salk, Marilyn Monroe, Mohandas Gandhi, Ruth Graham, Bill Gates, Agatha Christie, Nelson Mandela, Mick Jagger, your next-door neighbor, and your favorite professor. Clearly, there isn't one kind of life that all of these individuals should pursue. Different lives may be significant in different ways.

Despite this variety, however, it is easy to give a list of things that anyone might find meaningful:

- *Satisfying personal relationships.* Nothing in life contributes to individual happiness as much as loving people who love you back. That is why family and friends are so important.
- *Accomplishments one can be proud of.* You might take pride in what you achieve at school, at work, at home, in the gym, or in pursuit of your hobbies. If you're not proud of what you do, then it's hard to be happy, no matter how good your life is otherwise.
- *Aesthetic appreciation.* You can find beauty in many things, such as reading books, seeing *The Godfather*, playing over Bobby Fischer's chess games, listening to your favorite music, and watching a thunderstorm.
- *Enjoyable activities.* You might enjoy playing bridge, eating good food, shooting baskets, having sex, climbing mountains, traveling, quilting, or acting in a play.
- *Learning.* Learning satisfies one's curiosity and deepens one's understanding of the world. Ignorance is

rarely bliss; increasing one's knowledge is almost always satisfying.

- *Contributing to the welfare of others.* According to the studies, happy people are less self-absorbed than unhappy people and are more likely to help others in need. The studies also suggest that spending money on other people will make you happier than spending money on yourself. The ancients were right: Virtue does bring happiness.

There are many good things in life—so many, in fact, that it seems almost absurd for people to say that there's nothing worth living for.

Socrates's life was a combination of such things. He had a wife and children, and we are told that he was a faithful husband and father. His friends were devoted to him. He was a loyal Athenian who fought for his city in military campaigns. He enjoyed art and athletics. But above all, Socrates was dedicated to discovering the truth about the world. He lived in a time when not much was known, but his contribution to human understanding was immense. Through his teachings, Socrates opened up lines of thought that are still being explored today. And at the end of his life, he chose to die rather than betray the values he had lived for. It is easy to understand the meaning of his life—we need only reflect on all this.

However, it is also easy to be skeptical. In a certain frame of mind, people are attracted to the idea that really there is nothing to live for. They may concede that people *believe* various things to be worthwhile. But it is only an illusion. A hardheaded assessment, they say, exposes the truth: Life has no meaning.

Can the skeptic be answered? The only way to do so would be to consider the various things on our list, one at a time, and explain why each of them is meaningful. We can describe the ways in which personal relationships and accomplishments contribute to human happiness. We can expound on the glories of Mozart and the pleasures of basketball. We can observe that humans are by nature curious, and so we naturally want to understand the world. We can observe that human beings are social animals, so contributing to the welfare of others comes naturally to us.

Such reasoning may not show that our lives are "important to the universe," but it will accomplish something similar. It will show that we have good, objective reasons to live in some ways

rather than others. When we step outside our personal perspectives and consider humanity from an impersonal standpoint, we still find that human beings are the kinds of creatures who can enjoy life best by devoting themselves to such things as family and friends, work, music, mountain climbing, and all the rest. It would be foolish, then, for creatures like us to live in any other way.

How to Evaluate Arguments

Formulating and assessing arguments is important in every subject. However, it is especially important in philosophy. Philosophy often deals with big, abstract issues, and our only way to get a grip on those issues is through careful argumentation. A philosophical theory is only as good as the arguments that support it.

Some arguments are sound and some are not, and we need to know how to tell the difference. Unfortunately, there is no easy way to do this. Arguments come in many varieties, and they can go wrong in numerous ways. There are, however, some general principles that we should follow in evaluating arguments.

1. Ask the Most Important Questions.

The word “argument” often means a quarrel, and there is a hint of unpleasantness in the word. Children may be taught not to argue because it isn’t polite. In philosophy, however, an argument is simply a chain of reasoning designed to prove something. Arguments always consist of one or more premises and a conclusion. The conclusion is what we are trying to prove, and the premises are the considerations that are supposed to prove it.

Aristotle was the first person to study logic systematically. He gave this example, and teachers have been using it ever since:

- (1) All men are mortal.
- (2) Socrates is a man.
- (3) Therefore, Socrates is mortal.

This example illustrates the main features of an argument: There are two premises and a conclusion, and the conclusion follows from the premises.

What does it mean to say that the conclusion “follows from” the premises? In logic, this is not just a loose expression; it is a precise technical term. It means that *it would be impossible for the premises to be true and the conclusion to be false at the same time*. In Aristotle’s example, it is easy to see that the conclusion follows from the premises: It would be impossible for all men to be mortal and Socrates to be a man and yet for Socrates not to be mortal.

Sometimes, however, whether a conclusion follows from a particular set of premises is not obvious. Consider this argument:

- (1) Different cultures have different moral codes. For example, in some cultures, polygamy is thought to be all right, while in other cultures, it is thought to be wrong.
- (2) Therefore, there is no such thing as objective truth in ethics. Ethics is merely a matter of opinion, which varies from society to society.

This argument is discussed in chapter 11. Many people instinctively feel that the conclusion follows from the premises, but it does not. The fact that different cultures have different moral codes doesn’t imply that ethics is merely a matter of opinion. After all, it is still possible that some cultures are objectively right, and some are objectively wrong.

To mark these distinctions, logicians use a bit of terminology—they distinguish *validity* from *soundness*. An argument is valid if the conclusion follows from the premises. To be sound, however, an argument must be valid *and* its premises must be true. Sound arguments are what we aim for.

It is important to distinguish *whether the premises of an argument are true* and *whether the conclusion follows from them*. In Aristotle’s example, the premises are both true, and the conclusion follows from them. But here is a different example in which the premises are all true and yet the conclusion (which is also true) does not follow from them:

- (1) Socrates was executed when he was 70.
- (2) Socrates was executed in 399 B.C.
- (3) Therefore, Socrates once served in the Athenian army.

The conclusion does not follow because it is possible for Socrates to have been executed in 399 B.C. at the age of 70 and yet never to have served in the Athenian army.

Here is an example in which the premises and conclusion are all false but the conclusion does follow:

- (1) All the Sophists were dim-witted.
- (2) Socrates was a Sophist.
- (3) Therefore, Socrates was dim-witted.

One further point should be noted. The premises of an argument may support the conclusion more or less strongly. Sometimes the premises do not absolutely prove that the conclusion is true, but they provide good evidence for it. For example:

- (1) Lola was seen kissing Frankie a few minutes before she left the party.
- (2) After the party, Lola said she had a crush on a new guy.
- (3) Therefore, Lola has a crush on Frankie.

This argument does not absolutely prove that Lola has a crush on Frankie, because she might have kissed Frankie in a game of Spin the Bottle, or she might have been lying about having a crush on someone. But, given the truth of (1) and (2), she probably does have a crush on Frankie. The premises make it likely that the conclusion is true.

Thus, the most important advice for evaluating arguments is this: Start by asking whether the premises are true and whether the conclusion follows from them. If the conclusion doesn't follow, ask whether it is nevertheless supported by the premises.

2. Do Not Believe an Argument to Be Sound Merely Because You Agree with Its Conclusion.

People often make the mistake of liking an argument because they like its conclusion. The conclusion might be that God exists, or that abortion is acceptable, or that the government has imprisoned aliens in New Mexico. Whatever the conclusion might be, if you agree with it, then you may be biased in assessing the argument for it.

There are many unsound arguments with true conclusions. For example:

- (1) Everyone from Georgia has, at one time, been president.
- (2) Jimmy Carter is from Georgia.
- (3) Therefore, at one time, Jimmy Carter was president.

The conclusion is true, but the argument is unsound because its first premise is false. So you should never judge an argument by its conclusion. Whether the conclusion is true is one thing; whether the argument proves it, is another.

3. If an Argument Is Flawed, Try to Fix It.

If an argument turns out to be unsound, we should try to recast it into a better form. After all, we are interested in discovering the truth. An argument might fail while suggesting a better argument that succeeds.

A good example of this is Socrates's argument for obeying the law, which is discussed in chapter 1. Socrates is considering whether he should drink the deadly hemlock as ordered by the Athenian court. He gives this argument:

- (1) If we do not obey the law, then the state cannot exist.
- (2) It would be disastrous if the state didn't exist.
- (3) Therefore, we must obey the law.
- (4) The law has ordered me to drink the hemlock.
- (5) Therefore, I must drink the hemlock.

This argument has a flaw: The state can still exist even if we break the law once in a while. It is only necessary that we obey the law most of the time, as a general rule. So the argument, as stated, is unsound. That is not surprising, since it seems like there are cases in which we should ignore the law. Perhaps the law is blatantly unjust, or perhaps the circumstances are extremely unusual. Yet Socrates is clearly onto something. If we want to know why we should usually obey the law, this modified version of his argument seems right:

- (1) If we do not in general obey the law (allowing only rare exceptions), then the state cannot exist.

- (2) It would be disastrous if the state did not exist, because we would all be much worse off without it.
- (3) Therefore, we should in general obey the law (allowing only rare exceptions).

This version of the argument does not entail that Socrates had to drink the hemlock. Unfortunately, Socrates seems to have missed this point.

4. Watch for Common Fallacies.

Some mistakes are so common that they have been given names. Here are a few of them.

Equivocation. Equivocation is using a word in more than one sense. When the meaning of a word shifts midway through an argument, the argument is ruined. Mae West joked that “Marriage is a great institution, but I’m not ready for an institution.” If we took this joke to be a serious argument, we would get:

- (1) Marriage is a great institution.
- (2) Mae isn’t ready for an institution.
- (3) Therefore, Mae isn’t ready to be married.

Here the word “institution” is used in different ways. In the first premise, “institution” means something like “cultural tradition.” In the second premise, “institution” is short for “mental institution.” The argument equivocates.

Socrates appears to commit the fallacy of equivocation in Book I of Plato’s *Republic*. There Socrates is discussing the nature of justice with Polemarchus, who suggests that justice is helping one’s friends and harming one’s enemies. Socrates objects that it can never be right to harm anyone:

Socrates: But is it really right to harm any man?

Polemarchus: It certainly is, we ought to harm bad men who are our enemies.

Socrates: If we harm a horse do we make it better or worse?

Polemarchus: Worse.

Socrates: Worse, that is, by the standards by which we judge horses?

Polemarchus: Yes.

Socrates: And a dog if harmed becomes a worse dog by the standards of canine excellence?

Polemarchus: Surely.

Socrates: But must we not then say of a man that if harmed he becomes worse by the standards of human excellence?

Polemarchus: Certainly.

Socrates: But is not justice the standard of human excellence?

Polemarchus: It surely must be.

Socrates: Then will just men use their justice to make others not just? Or, in short, will good men use their goodness to make others bad?

Polemarchus: That cannot be so.

Socrates: Then since the just man is good, Polemarchus, it is not the function of the just man to harm either his friends or anyone else, but of his opposite, the unjust man.

Polemarchus: What you say is perfectly true, Socrates.

The gist of Socrates's argument is this:

- (1) To harm someone is to make him worse.
- (2) People are better or worse according to whether they are just or unjust.
- (3) Therefore, to harm someone is to make him more unjust.
- (4) It is never right to make someone more unjust.
- (5) Therefore, it is never right to harm someone.

The problem is that the word "worse" is used in two senses. You can be made *worse off* in many different ways: if you get sick, if you lose all your money, if your friends turn on you, and so on. But in another sense, you become *morally* worse—you become a worse person—only if you do something bad or if you undergo a change in character. Socrates's argument trades on this ambiguity. The first premise appears to be true because we have the first meaning in mind—to harm someone is to make him worse off—while the second premise appeals to the second meaning—people are morally better or worse according to whether they're just or unjust. But if we choose one meaning and stick to it, then one or the other of the premises turns out to be false.

Appeal to Authority. When people have no good arguments for their views, they will sometimes appeal to an authority to prove a point. Such appeals should always be viewed with suspicion. We may always ask why the so-called authority says what he says, and if there is no good answer, then the authority may be ignored.

Often, of course, there is nothing wrong with consulting an authority to get information. If you want to know Wade Boggs's batting average for 1988, you can look it up in *The Baseball Encyclopedia*. Or if you want to know what movies Charlize Theron has been in, checking the Internet Movie Database (imbd.com) will work fine.

Ultimately, however, what justifies us in believing that Wade Boggs batted .366 in 1988 is not that *The Baseball Encyclopedia* says so. We trust *The Baseball Encyclopedia* because we believe that somebody kept up with Boggs's hits and at-bats and did the required calculation. That work justifies our saying that Boggs batted .366. Similarly, we believe that Charlize Theron was in *Monster* because the Internet Movie Database would not be so widely used if it were not usually accurate. The mere fact that an authority has said something is never the last word.

It is easy to overlook this point when ethics and religion are involved. Some people argue:

- (1) Leviticus says homosexuality is an abomination.
- (2) Therefore, homosexuality is an abomination.

But what lies behind this judgment? Why does Leviticus say that homosexuality is an abomination? We need to focus on the reasons behind the pronouncement, rather than on the pronouncement itself. If the authority is reliable, then there must be good reasons for what it says. But merely citing the authority does not indicate what those reasons are. Thus, if homosexuality is an abomination, there must be some other argument that proves it. If there is none, then the authority's pronouncement is unfounded. This is not just a point about religious authority but about authority in general.

Ad Hominem. "Ad hominem" is Latin for "against the man"; it is the fallacy of trying to discredit an opinion by discrediting

those who hold it. An *ad hominem* argument is thus the reverse of an appeal to authority. Consider:

- (1) Charles Darwin was an atheist.
- (2) Therefore, his godless theory should be ignored.
- (1) Martha Stewart was convicted of obstruction of justice.
- (2) Therefore, we can ignore what she says about cooking.

The personal remarks are irrelevant. Darwin and Stewart both have reasons for what they say, and if we are trying to assess their views, it is the strength of their arguments, and not their personal histories or characters, that matters.

Begging the Question. In everyday speech, to “beg the question” is just to say something that brings a certain question to mind. For example, “I heard that Sophie’s bank account is running low. This begs the question: What did she do with her parents’ money?” But in philosophy, an argument “begs the question” if its conclusion is smuggled in among its premises. Here, for example, is a question-begging argument in defense of free will:

- (1) People control their own destinies—they can choose how they will behave and what sorts of individuals they will be.
- (2) Therefore, people have free will.

In this argument, the very thing we are trying to prove—that people have free will—is asserted (in different words) in the first premise. So the argument does not prove that people have free will; instead, the argument simply assumes it. Here is another example:

- (1) Capital punishment is murder.
- (2) Murder is wrong.
- (3) Therefore, capital punishment is wrong.

This argument begs the question because of premise one: “Capital punishment is murder.” If you think that a type of killing is justified, then you don’t call it murder. For example, you would say, “He killed the intruder in self-defense,” not “He murdered the intruder in self-defense.” “Murder,” by definition, is wrongful killing. Thus, the first premise is saying “Capital

punishment is wrongful killing,” which is the very thing the argument is trying to prove.

Begging the question is related to another fallacy known as *circular reasoning*. The simplest example of circular reasoning is when you give A as your evidence for B and then turn around and give B as your evidence for A—for example, “I know that God exists because the Bible says so. How do I know the Bible can be trusted? Because it is the word of God.”

False Dilemma. An argument presents a false dilemma if it says that we must choose between a limited number of options, when really there are more options available. Here is a version of an argument that we discussed in chapter 2:

- (1) The human eye arose either by chance or by intelligent design.
- (2) It could not have arisen by chance.
- (3) Therefore, it is the product of intelligent design.

The problem is that there are at least three ways of accounting for the eye: chance, intelligent design, and evolution by natural selection. So, the first premise of the argument is false.

5. Don't Reject an Idea Merely Because You've Heard a Bad Argument for It.

One final piece of advice: Don't reject a thesis merely because one particular argument for it is unsound. The failure of an argument does *not* mean that the conclusion of the argument is false; it means only that *this particular argument* does not prove that conclusion to be true. A different argument might succeed. In other words, don't blame the conclusion just because someone has thought up a bad argument for it; it's not the conclusion's fault.

Notes on Sources

Chapter 1: The Legacy of Socrates

“pushing them on the streets of Athens”: Gregory Vlastos, *Socrates: Ironist and Moral Philosopher* (Cambridge University Press, 1991), p. 294.

“Suppose that while we were preparing to run away . . .” Plato, *Crito*, tr. Hugh Tredennick, *The Last Days of Socrates* (Harmondsworth, Middlesex: Penguin, 1969), p. 89.

“Then since you have been born . . .” Plato, *Crito*, pp. 90–91.

“[A]ny Athenian, on attaining to manhood . . .” Plato, *Crito*, p. 92.

“He is absolutely unique . . .” Plato, *Symposium*, tr. Michael Joyce, *The Collected Dialogues of Plato*, ed. by Edith Hamilton and Huntington Cairns (Princeton University Press, 1961), p. 572.

Chapter 2: God and the Origin of the Universe

The top-of-the-chapter quote is from the couplet: “The universe embarrasses me, and I cannot think/That this watch exists and has no watchmaker.” Voltaire, *Les Cabales* (1772), lines 111–112, trans. by J. M. A. Perot in his *Man and God* (London: Freethought, 1881), p. 125.

The “recent Gallup poll” is from 2009: www.gallup.com/poll/1690/Religion.aspx#1. The international poll is from the Pew Global Attitudes Project: Spring 2005 Survey. The quote about American attitudes being similar to attitudes in developing nations is from the Pew Global Attitudes Project, “Among Wealthy Nations . . . U.S. Stands Alone in its Embrace of Religion” (released December 19, 2002).

The Gallup International Millennium Survey: www.gallup-international.com/ContentFiles/millennium15.asp.

“[I]ts several parts . . . are so formed and adjusted . . .” William Paley, *Natural Theology: or, Evidences of the Existence and Attributes of the Deity* (London: Faulder, 1802); repr. in *The Cosmological Arguments*, ed. by Donald R. Burrill (Garden City, NY: Anchor, 1967), p. 166.

“I was charmed and convinced . . .” *The Autobiography of Charles Darwin*, ed. by Nora Barlow (New York: Norton, 1958), p. 59. The next quotation is from p. 87.

"In living bodies, variation will cause . . ." Charles Darwin, *On the Origin of Species* (London: John Murray, 1859), p. 189.

"I can remember having a picture of the cosmos . . ." Peter van Inwagen, *The Possibility of Resurrection and Other Essays in Christian Apologetics* (Boulder, CO: Westview Press, 1998), p. 5.

"I can still call the image to mind . . ." van Inwagen, *Possibility of Resurrection*, p. 5.

Anselm, Gaunilo, and Kant on the Ontological Argument: The relevant writings are in *The Cosmological Arguments*, ed. by Donald R. Burrill (Garden City, NY: Anchor, 1967).

Chapter 3: The Problem of Evil

All quotations from the Book of Job are from chs. 8 and 11 of *The Holy Bible, Revised Standard Version* (New York: Thomas Nelson, 1952).

Plantinga's view: Alvin Plantinga, *God, Freedom, and Evil* (Grand Rapids, MI: Eerdmans, 1978).

"Every human being has an eternal future . . ." Peter van Inwagen, *God, Knowledge, and Mystery: Essays in Philosophical Theology* (Ithaca, NY: Cornell University Press, 1995), pp. 101–102.

"[A] man who drove a cement mixer truck . . ." Alvin Plantinga, "Self-Profile," in *Alvin Plantinga*, ed. by Peter van Inwagen and James E. Tomberlin (Dordrecht: Reidel, 1985), p. 34, quoted in Daniel Howard-Snyder, "God, Evil, and Suffering," *Reason for the Hope Within*, ed. by Michael J. Murray (Grand Rapids, MI: Eerdmans, 1999), p. 77.

"That there is much suffering in the world . . ." Darwin, *Autobiography*, p. 90.

A Note about the Book of Job. Chapter 3 begins with the story of Job, but readers familiar with the Bible might have noticed that I did not mention the beginning or end of the story. The reason is that modern scholarship tells us that the Book of Job as we have it today consists of two parts, an original core and a "frame"—an introduction and conclusion—added later by unknown individuals. In the original story, we meet Job, we learn of his suffering, and we hear the accusations of his friends. But we are assured that Job is blameless. That is why his sufferings are a mystery. The mystery is deepened when Job observes that other, less virtuous men suffer no such misfortunes. The original story ends there, as a mystery that has no answer.

The frame eliminates the mystery and changes a profound parable into an exasperating story about a capricious God. In the introduction that was added, an explanation is given of why God permitted Job's suffering. One day, we are told, when God's angels had gathered around, God singled out Job for special praise. But Satan scoffed and

said that Job was God's loyal servant only because God had rewarded him with riches and a loving family. To prove Satan wrong, God allowed him to take away all that Job had. Thus, God permitted Satan to torment Job simply to prove a point: He wanted to demonstrate that Job would remain faithful to him no matter what.

In this version of the story, God is not the just ruler that Job's friends believed him to be. Instead, he is no better than a human father who would allow his children to be tormented in order to prove to a skeptic that they would still love him. The tacked-on conclusion makes things even worse. In an apparent attempt to vindicate God, he is shown restoring Job's prosperity. This does not really help, though, for the terrible, unmerited suffering that Job has already endured cannot be wiped out. It remains forever a part of his history.

Chapter 4: Do We Survive Death?

The full quotation at the top of the chapter is: "Then soules and bodies have one birth and growth, And the same wastes of age decay them both." Lucretius, *Lucretius: De Rerum Natura* (London: Duckworth, 1996; originally ca. 50 B.C.), tr. Lucy Hutchinson, ed. by Hugh de Quehen, p. 96, lines 445–446.

"By death do we not mean . . ." Plato, *Phaedo*, tr. R. Hackforth (Indianapolis: Bobbs-Merrill, 1955), p. 44.

"Your view about the soul . . ." Plato, *Phaedo*, p. 58.

Nearly half of Americans believe in ESP, etc.: This poll was conducted October 16–18, 2007, by the Associated Press and Ipsos. I read about it on cnn.com: "Poll: Third of People Believe in Ghosts," October 26, 2007 (an Associated Press story with no byline).

Raymond A. Moody, Jr., *Life after Life: The Investigation of a Phenomenon—Survival of Bodily Death*, 25th Anniversary Edition (New York: HarperSanFrancisco, 2001; originally 1975). The quotations are from pp. 169 and xxv–xxvi.

Moody coined the phrase, "near-death experience": Sanjay Gupta, *Cheating Death* (New York: Wellness Central, 2009), p. 98.

Craig Hamilton-Parker, *What to Do When You Are Dead: Living Better in the Afterlife* (New York: Sterling, 2001).

Today there is a "field" of near-death research, with its own professional association and its own journal: www.iands.org/pubs/jnds.

"Of male survivors of cardiac arrest . . ." Quoted by Paul Edwards in *Reincarnation: A Critical Examination* (Amherst, NY: Prometheus Books, 1996), p. 150.

Researchers can induce out-of-body experiences by stimulating the superior temporal gyrus: Dirk de Ridder *et al.*, "Visualizing

out-of-body experience in the brain," *The New England Journal of Medicine*, vol. 357, no. 18 (2007), pp. 1829–1833. People who achieve out-of-body experiences through prayer and meditation show reduced activity in their superior parietal lobes: Gupta, *Cheating Death*, pp. 110–111.

One psychiatrist compared near-death experiences with drug-induced hallucinations: Ronald K. Siegel, "Life after Death," ed. by George O. Abell and Barry Singer, *Science and the Paranormal* (New York: Scribner, 1981), pp. 159–184.

On Nelson's hypothesis that near-death experiences are manifestations of REM sleep, see Gupta, *Cheating Death*, pp. 111–133.

The quotations from A. J. Ayer are from his *The Meaning of Life* (New York: Scribner, 1990), pp. 200, 201, 204.

"It is little less than scandalous . . ." Edwards, *Reincarnation*, pp. 223–224.

"an Ireland that never was . . ." Quoted by Martin Gardner in *Fads and Fallacies in the Name of Science* (New York: Dover, 1957), p. 316.

Champe Ransom's critique of the Stevenson studies is quoted in Edwards, *Reincarnation*, pp. 276–277.

"I should be willing now to stake . . ." William James, *Proceedings of the Society for Psychical Research*, vol. 6 (1889–1890), p. 654, quoted in Martin Gardner, *On the Wild Side* (Buffalo, NY: Prometheus Books, 1992), p. 217.

For information about cold reading, I am indebted to Joe Nickell, "John Edward: Hustling the Bereaved," *Skeptical Inquirer*, November/December 2001, pp. 19–22.

"[m]ore than 6,000 people were asked . . ." Richard Wiseman and Ciaran O'Keefe, "A Critique of Schwartz et al.'s After-Death Communication Studies," *Skeptical Inquirer*, November/December 2001, p. 28.

"They're telling me to acknowledge Anthony." Nickell, "John Edward: Hustling the Bereaved," p. 19.

"When anyone tells me that he saw a dead man restored to life . . ." David Hume, *An Enquiry Concerning Human Understanding*, ed. by Charles W. Hendel (New York: Liberal Arts Press, 1955; originally 1748), pp. 123–124.

"We frequently hesitate concerning the reports of others." Hume, *Enquiry*, p. 120.

"an everlasting check . . ." Hume, *Enquiry*, p. 118.

Chapter 5: The Problem of Personal Identity

The top-of-the-chapter quote is from Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), p. 33.

The quotations from David Hume about the Bundle Theory are from *A Treatise of Human Nature* (1739), I, IV, 6.

“a conscious system in its own right . . .” Roger Sperry, quoted by Norman H. Horowitz in *Roger Wolcott Sperry*, published by the Nobel Foundation, Stockholm, 1999.

Derek Parfit’s argument about the split-brain cases and the Bundle Theory is from his essay “Divided Minds and the Nature of Persons,” ed. by Colin Blakemore and Susan Greenfield, *Mindwaves* (London: Blackwell, 1987), pp. 19–28.

Plutarch discusses the Ship of Theseus in *Vita Thesei* (The Life of Theseus), pp. 22–23.

The information about cell turnover was gleaned mostly from Nicholas Wade, “Your Body Is Younger Than You Think,” *The New York Times*, August 2, 2005.

“Everyone sees he would be the same person . . .” John Locke, *An Essay Concerning Human Understanding* (1690), II, xxvii, 15. I have embellished Locke’s example without changing its point.

“Suppose a brave officer . . .” Thomas Reid, *Essays on the Intellectual Powers of Man* (1785), III, 6.

The study of people’s memories of the *Challenger* disaster is Ulric Neisser and Nicole Harsch, “Phantom Flashbulbs: False Recollections of Hearing the News about *Challenger*,” E. Winograd and U. Neisser, eds., *Affect and Accuracy in Recall: Studies of “Flashbulb Memories”* (New York: Cambridge University Press, 1992), pp. 9–31.

“[O]ne should really think it self-evident . . .” Joseph Butler, *The Analogy of Religion* (1736), Dissertation I: Of Personal Identity.

Derek Parfit discusses how a belief in the Bundle Theory can be liberating in *Reasons and Persons* (Oxford University Press, 1984), pp. 281–282.

Chapter 6: Body and Mind

Elizabeth noted that, for a body to be moved, it must be pushed: See Princess Elizabeth, Letter to Descartes, May 6–16, 1643, in René Descartes, *Philosophical Writings*, tr. Elizabeth Anscombe and Peter Geach (Indianapolis: Bobbs-Merrill, 1971), pp. 274–275.

“I cannot see why this should convince us . . .” Princess Elizabeth, Letter to Descartes, June 10–20, 1643, in Descartes, *Philosophical Writings*, p. 278.

Elizabeth complained that immaterial substances are always defined in terms of what they are not: Princess Elizabeth, Letter to Descartes, June 10–20, 1643, in Descartes, *Philosophical Writings*, p. 278.

“the dogma of the Ghost in the Machine”: Gilbert Ryle, *The Concept of Mind* (London: Hutchinson, 1949), pp. 15–16.

Wilder Penfield’s work is described in his book *The Excitable Cortex in Conscious Man* (Springfield, IL: Charles C Thomas, 1958).

The example about beer cans is from John R. Searle, "The Myth of the Computer," *New York Review*, April 29, 1982, pp. 3–5.

"what it is like": Thomas Nagel, "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (1974), pp. 435–450. Nagel introduced this phrase as a way of referring to the subjective dimension of consciousness.

Frank Jackson, "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982), pp. 127–136.

Chapter 7: Could a Machine Think?

"It is indeed conceivable . . ." René Descartes, *Discourse on Method*, tr. Laurence J. Lafleur (Indianapolis: Bobbs-Merrill, 1960; originally 1637), pp. 41–42. I have added a comma after the word "presence."

For more on Watson, see IBM's website, www.ibm.com.

Alan Turing, "On Computable Numbers, with an Application to the *Entscheidungsproblem*," *Proceedings of the London Mathematical Society*, series 2, 42 (1936–1937), pp. 230–265.

"I won't say that what Turing did made us win the war . . ." I. J. Good, quoted in Jack Copeland, *Artificial Intelligence: A Philosophical Introduction* (Oxford: Blackwell, 1993), p. 10.

Alan Turing, "Computing Machinery and Intelligence," *Mind* 59 (1950), pp. 433–460.

"If trees could converse with us . . ." Copeland, *Artificial Intelligence*, p. 38.

Eliza on the internet: www.chayden.net/eliza/Eliza.html.

Daniel Bobrow, "A Turing Test Passed," *ACM SIGART Newsletter*, December 1968, pp. 14–15; quoted in Copeland, *Artificial Intelligence*, p. 39.

For information on the Loebner Prize: www.loebner.net/Prizef/loebner-prize.html.

The Chinese Room Argument was first presented in John R. Searle, "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3 (1980), pp. 417–424.

Chapter 8: The Case against Free Will

All quotations from Clarence Darrow are from *Attorney for the Damned: Clarence Darrow in the Courtroom*, ed. by Arthur Weinberg (University of Chicago Press, 1957): "I really do not in the least believe in crime" is from p. 3. The headline from the *Chicago Evening Standard* is from p. 19. "Intelligent people now know" is from p. 56. "I do not know what it was" is from p. 37. "Is Dickie Loeb to blame" is from p. 55. "Nature is strong and she is pitiless" is from pp. 64–65.

Laplace's *Philosophical Essay on Probability* was published in 1819. The relevant portions are translated by John Cottingham in *Western Philosophy: An Anthology*, ed. by John Cottingham (Oxford: Blackwell, 1996).

Wilder Penfield, *The Excitable Cortex in Conscious Man* (Springfield, IL: Charles C Thomas, 1958).

The story of the patient who heard Guns n' Roses when his brain was stimulated is told by Daniel Dennett in *Consciousness Explained* (Boston: Little, Brown, 1991), pp. 58–59.

"head turning and slow displacement . . ." Jose M. R. Delgado, *Physical Control of the Mind* (New York: Harper, 1969), p. 115.

For information about the Kornhuber experiment, see Benjamin Libet, "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action," *The Behavioral and Brain Sciences* 8 (1985), pp. 529–566.

Differences in what we name our children: Steven D. Levitt and Stephen J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything* (New York: HarperCollins, 2005), pp. 184–185, 192, 202. The authors rely on Roland G. Fryer Jr. and Steven D. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, vol. 119, no. 3 (August 2004), pp. 767–805.

2+ million Americans in jail or prison; 5 million on probation or parole: <http://bjs.ojp.usdoj.gov/content/glance/tables/corr2tab.cfm>.

Social factors predictive of criminal behavior: Levitt and Dubner, *Freakonomics*, pp. 138–139. The authors give sources for their claims on pp. 222–223.

Information about Eric Rudolph: *Newsweek*, July 27, 1998, pp. 16ff.

"If we are to use the methods of science . . ." B. F. Skinner, *Science and Human Behavior* (New York: Free Press, 1953), p. 6.

The guards-and-prisoners experiment: Philip Zimbardo, "The Pathology of Imprisonment," *Society* 9 (1972), pp. 4–8.

The Milgram experiment: Stanley Milgram, *Obedience to Authority* (New York: Harper and Row, 1974).

"And who is my neighbor? . . ." Luke 10:29–37, *The Holy Bible, Revised Standard Version* (New York: Thomas Nelson, 1952).

The Good Samaritan experiment: J. M. Darley and C. D. Batson, "'From Jerusalem to Jericho': A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology* 27 (1973), pp. 100–108.

The information about twins is from Lawrence Wright, *Twins: And What They Tell Us about Who We Are* (New York: John Wiley, 1997).

The information about biology and violence is from William Wright, *Born That Way: Genes, Behavior, Personality* (New York: Routledge, 1999).

“Though we no longer say . . .” Judith Rich Harris, *The Nurture Assumption* (New York: Free Press, 1998), pp. 295–296.

The quotation about orchid children is from David Dobbs, “The Science of Success,” www.theatlantic.com/magazine/archive/2009/12/the-science-of-success/7761, p. 4 of 14. This essay appeared as “Orchid Children” in the print version (*The Atlantic*, December 2009).

Chapter 9: The Debate over Free Will

The quotations from Dr. Johnson are from Boswell’s *The Life of Samuel Johnson, LL.D.* (1791), section 273.

I have given Isaac Bashevis Singer’s dates as 1904–1991. However, some respectable sources say that he was born in 1902; others say 1908.

“If Determinism is true . . .” Peter van Inwagen, *An Essay on Free Will* (Oxford University Press, 1983), p. 16.

“No man treats a motorcar . . .” Bertrand Russell, “Has Religion Made Useful Contributions to Civilization?” (1930), repr. in *Why I Am Not a Christian* (New York: Simon and Schuster, 1957), p. 40.

Information about Eric Rudolph: *Newsweek*, July 27, 1998, pp. 16ff.

Chapter 10: Our Knowledge of the World around Us

The top-of-the-chapter quote is from Woody Allen, “On Seeing a Tree in Summer,” *The Insanity Defense* (New York: Random House, 2007), p. 150.

“there are no conclusive indications . . .” René Descartes, *Meditations Concerning First Philosophy*, tr. Laurence J. Lafleur, in Descartes, *Philosophical Essays* (Indianapolis: Bobbs-Merrill, 1964), p. 77.

“though it were possible . . .” Berkeley, *A Treatise Concerning the Principles of Human Knowledge* (Indianapolis: Bobbs-Merrill, 1957; originally 1710), p. 31 (paragraph 18).

Bertrand Russell’s phrase “the advantages of theft over honest toil” is from his *Introduction to Mathematical Philosophy* (London: Allen and Unwin, 1919), p. 71.

“When [organisms] apprehend the world . . .” Steven Pinker, *How the Mind Works* (New York: Norton, 1997), p. 212. My summary of how vision works is based on Pinker’s discussion.

“Objects have regular, compact silhouettes . . .” Pinker, *How the Mind Works*, p. 217.

Irving Biederman introduced his theory of geons in “Recognition-by-Components: A Theory of Human Image Understanding,” *Psychological Review* 94 (1987), pp. 115–147.

On Jonathan I., see Oliver Sacks, *An Anthropologist on Mars: Seven Paradoxical Tales* (New York: Knopf, 1995), pp. 3–41 (quotation on p. 7; also see p. 15). At first, Jonathan I. thought his experience was like watching a black-and-white television (p. 3), but later he retracted that description (p. 10). He also denied that the word “grey” accurately described what he saw—rather, he said, ordinary language had no words for his experiences (p. 11). Eventually, Jonathan I. grew accustomed to his world, relishing its aesthetic properties as well as his extraordinary night vision (pp. 37–39). When given a chance to try to regain his color vision, he found the suggestion to be unintelligible and repugnant (p. 39).

Eighty percent of the fibers going to the vision center come from the brain: Atul Gawande, “The Itch,” *The New Yorker*, June 30, 2008, pp. 58–65. This fact is mentioned on p. 63. Also on p. 63 is the quotation: “The account of perception that’s starting to emerge . . .”

Chapter 11: Ethics and Objectivity

“While we had been talking . . .” Plato, *The Republic*, tr. H. D. P. Lee (Harmondsworth, Middlesex: Penguin, 1955), p. 63.

“The life of a man . . .” David Hume, *Essays: Moral, Political, and Literary* (Oxford University Press, 1963; originally 1741–1742), p. 590.

In Islamic countries, men may have more than one wife: See, for example, J. Patrick Gray, “A Corrected *Ethnographic Atlas*,” *World Cultures*, vol. 10, no. 1 (1999), pp. 24–85.

The example of the Pakistani girl is from *Newsweek*, July 15, 2002, p. 18.

The example of the Nigerian woman is from *The New York Times*, August 20, 2002, p. A12.

Sharia has been adopted by 12 of Nigeria’s 36 states: Karin Brulliard, “In Nigeria, Sharia Fails to Deliver: Vows of Accountability, Equity Under Moderate Islamic Law Are Unmet,” *The Washington Post*, August 12, 2009.

“Take any action allow’d to be vicious . . .” David Hume, *A Treatise of Human Nature* (Oxford University Press, 1888; originally 1739), p. 468.

“There are no moral phenomena . . .” Friedrich Nietzsche, *Beyond Good and Evil*, tr. Walter Kaufmann (New York: Vintage Books, 1966; originally 1886), p. 108.

Chapter 12: Why Should We Be Moral?

The top-of-the-chapter quote is from H. L. Mencken, *A Mencken Chrestomathy* (New York: Knopf, 1956), p. 617.

Glaucon's account of how Gyges became king is in Book I of Plato's *Republic*.

Homosexuality is condemned in Leviticus 18:22. Washing your clothes after touching a menstruating woman's chair is required in Leviticus 15:22. Bearing false witness is condemned in Exodus 20:16.

"In all determinations of morality . . ." David Hume, *An Inquiry Concerning the Principles of Morals* (Indianapolis: Bobbs-Merrill, 1957; originally 1752), pp. 12–13.

"[T]he question depends upon what idea we have formed of the moral character of the Deity. . . ." John Stuart Mill, *Utilitarianism* (Indianapolis: Bobbs-Merrill, 1967; originally 1861), p. 28.

"I feel that I am bound . . ." Mill, *Utilitarianism*, p. 34.

"The firm foundation . . ." Mill, *Utilitarianism*, p. 40.

8.8 million children under the age of 5 die each year, mostly from preventable causes: Robert E. Black *et al.*, "Global, Regional, and National Causes of Child Mortality in 2008: A Systematic Analysis," *The Lancet*, vol. 375, no. 9730 (June 5, 2010), pp. 1969–1987.

"go very far out of our way . . ." Jan Narveson, *Moral Matters* (Peterborough, Ontario: Broadview Press, 1993), p. 131.

"What about parting with the means . . ." Narveson, *Moral Matters*, p. 145.

"Reason makes it possible . . ." Peter Singer, *How Are We to Live?* (Amherst, NY: Prometheus Books, 1995), p. 229.

Chapter 13: The Meaning of Life

"I had what might truly be called an object in life . . ." John Stuart Mill, *Autobiography* (1873), in *Essential Works of John Stuart Mill*, ed. by Max Lerner (New York: Bantam Books, 1961), p. 83.

"I was in a dull state of nerves . . ." Mill, *Autobiography*, p. 83.

"Relieved from my ever present sense . . ." Mill, *Autobiography*, pp. 87–88.

The phrase "the point of view of the universe" comes from Henry Sidgwick, *The Methods of Ethics* (London: Macmillan, 1874), p. 382.

"In seeing ourselves from outside . . ." Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), p. 214.

"Ask not that events should happen as you will . . ." *The Manual of Epictetus*, tr. P. E. Matheson, in *The Stoic and Epicurean Philosophers*, ed. by Whitney J. Oates (New York: Random House, 1940), p. 470. The full sentence is: "Ask not that events should happen as you will, but let your will be that events should happen as they do, and you will have peace."

The elderly tend to be happier than people in their 20s and 30s: Lindsey Tanner, "Gray Citizens More Apt to Be Happy: Later

Life Perceived as Better, Study Finds,” *The Associated Press*, published in *The Birmingham News*, April 19, 2008, and Nicholas Bakalar, “Happiness May Come with Age, Study Says,” *The New York Times*, May 31, 2010.

Men tend to be happier than women: David Leonhardt, “He’s Happier, She’s Less So,” *The New York Times*, September 26, 2007.

Eighty percent of the variation in happiness is due to genes: The study by Auke Tellegen and David Lykken (1996) is mentioned on p. 40 of Michael Wiederman, “Why It’s So Hard to Be Happy,” *Scientific American Mind* (February/March 2007), pp. 36–43.

People in wealthy countries are happier than people in poor countries: David Leonhardt, “Maybe Money Does Buy Happiness After All,” *The New York Times*, April 16, 2008.

Americans are richer than Mexicans but less happy: from a survey conducted in 2007 by the Pew Global Attitudes Project. See p. 2 of Bruce Stokes, “Happiness Is Increasing in Many Countries—But Why?” at <http://pewglobal.org/2007/07/24/happiness-is-increasing-in-many-countries-but-why/>.

Americans have gotten richer but not happier: from David G. Myers’s analysis of data from the 2000 U.S. Census and the National Opinion Research Center, as reported by Wiederman, “Why It’s So Hard to Be Happy,” p. 38.

Happiness is correlated with other things: David G. Myers, *The Pursuit of Happiness* (New York: Avon Books, 1992).

“Ask yourself if you are happy . . .” Mill, *Autobiography*, p. 88.

“Death is nothing to us.” Epicurus, *Letter to Menoeceus*, tr. C. Bailey, in Oates, *The Stoic and Epicurean Philosophers*, p. 30.

“Where are now the warring kings . . .” W. B. Yeats, *Song of the Happy Shepherd* (1889).

The quotation by Julius Henry “Groucho” Marx is from Fred R. Shapiro, *The Yale Book of Quotations* (New Haven: Yale University Press, 2006), p. 498.

Spending your money on other people will make you happier than spending it on yourself: Elizabeth W. Dunn, Lara B. Aknin, and Michael I. Norton, “Spending Money on Others Promotes Happiness,” *Science* 21, vol. 319. no. 5870 (March 2008), pp. 1687–1688.

Appendix: How to Evaluate Arguments

The discussion of justice between Socrates and Polemarchus is from Plato’s *Republic*, tr. H. D. P. Lee (Baltimore: Penguin Books, 1955), p. 61.

Index

A

Abortion, 144–146
Aboutness, 81–82
Accidents and free will, 121
Accomplishments, 173
Accountability, 114–115
Actions vs. movements, 98–99
Ad hominem arguments, 182–183
Aesthetic appreciation, 173
Afterlife
 evidence of, 41–49
 Hume’s argument against
 miracles, 49–51
 immortal soul, 38–41
 meaning of life and, 169–170
 personhood and, 53
 religion and indifferent
 universe, 172
 Same-Body Theory and, 59

AI (artificial intelligence). *See*
 Artificial intelligence (AI)
Alcibiades, 9
Alter, Torin, xii
Altruism, 159–164
Amazing Randi, x, 48
America. *See* United States
Analogy between state and
 one’s parents, 5–6
Analytical Behaviorism, 72–74.
 See also Behaviorism
Animals
 suffering of, 37
 values and, 149
Appeal to authority, 182
Arguments
 for AI, 84–87
Argument from Design,
 11–15

- Argument from
 - Disagreement, 144–146
- Argument from Lack of Proof, 146–148
- Chinese Room Argument, 91–93
- Conceivability Argument, 68
- cultural relativism, 140–144
- Descartes' response to skepticism, 130–133
- Determinist Argument, 109–110
 - evaluating, 176–184
 - for existence of God, 26
- First Cause Argument, 21–23
- God as necessary being, 23–25
- Hume's argument against miracles, 49–51
- Libertarian response to Determinist, 110–115
- Metaphysical Argument, 148–151
- of Socrates, 4–9
- Aristotle
 - evaluating arguments, 176
 - Fatalism, 97
- Artificial intelligence (AI)
 - argument for, 84–87
 - brains and computers, 83–84
 - free will and, 119
 - Functionalism and, 79
 - Turing Test, 87–91
 - why Turing Test fails, 91–93
- Athens
 - Socrates's condemnation, 1–3
 - Socrates's death, 3–9
- Austin, J. L., 134
- Authority, appeal to, 182
- Ayer, A. J., 44
- B**
- Badness. *See also* Morality
 - evaluating people's, 120
 - Ring of Gyges, 153–154
- Baseball Encyclopedia, The*, 182
- Batson, C. D., 103–104
- Begging the question, 183–184
- Behavior
 - genes and free will, 104–108
 - Social Contract Theory, 157–159
- Behaviorism
 - defined, 71–73
 - Determinism and, 101–102
 - failure of, 73–74
 - Turing Test, 91
- Behe, Michael J., 19
- Belief
 - accountability and free will, 114–115
 - meaning of life, 174
 - in mental states, 68–69
 - reasonableness of God, 10–11
 - religion and indifferent universe, 172
- Benevolence, 159–164
- Bentham, Jeremy
 - meaning of life, 166
 - Utilitarianism, 160
- Benthamites, 160
- Berkeley, George, 128–129
- Bernstein, Morey, 45
- Best-Explanation Argument
 - Darwin and, 20
 - defined, 12
 - Hume's objections to, 14–15
- Bible*
 - appeal to authority, 182
 - ethics and religion, 156
 - idea of immortal soul, 39–40
 - notes on Job, 186–187

- Biederman, Irving, 136
- Big Bang, 21–23
- Biology
- cell regeneration, 60
 - complexity in, 18–20
 - violence and, 106–108
- Blame
- morality and free will, 123–124
 - responsibility and free will, 120–123
- Bletchley Park, 87
- Body
- continuity of personal identity, 58–61
 - warning system, 30–31
- Body and mind
- Descartes and Elizabeth, 67–71
 - doubts about materialist theories, 80–82
 - materialist theories of mind, 71–80
- Boggs, Wade, 182
- Book of Job*, 27–29
- Brain
- activity in near-death experiences, 43
 - computers and, 83–84
 - Determinism, 98–99
 - Mind–Brain Identity Theory, 74–75
 - Piecemeal-Replacement Argument, 84–85
 - problem of reality, 125–127
 - Same-Brain Theory, 60–61
 - split brains, 55–57
 - vision and, 134–137
- Brentano, Franz, 81
- Bundle Theory
- conclusion, 65–66
 - defined, 53–55
 - personhood over time, 57–58
 - split brains and, 55–57
- Butler, Bishop Joseph, 65
- C**
- Capital punishment, 144–145
- Catastrophism, 16
- Cause and effect
- belief in God and, 14
 - Compatibilism and, 117
 - Determinism, 97–99
 - First Cause Argument, 21–23
- Cebes, 39
- Cell regeneration, 60
- Chain of memories, 62–63
- Character, moral. *See* Morality
- Chicago American*, 46
- Children and reincarnation, 46–47
- Chinese Room Argument, 91–93
- Choice. *See* Free will
- Christianity. *See also* Religion
- idea of immortal soul, 39–40
 - indifferent universe and, 171–173
- Circular reasoning, 184
- Cobbler and the prince, 59
- Coercion, 121
- Cogito, ergo sum*, 131
- Cold reading, 48–49
- Color perception, 136
- Common fallacies, 180–184
- Commonsense view of
- perception, 137–138
- Communication with the dead, 47–49
- Compatibilism, 115–118
- Computers
- artificial intelligence. *See* Artificial intelligence (AI)

- brains and, 83–84
- Chinese Room Argument, 91–93
- quantum mechanics and, 112–113
- Conceivability Argument
 - defined, 68
 - vs. Mind–Brain Identity Theory, 75
- Concept of Mind, The* (Ryle), 72
- Conclusions
 - begging the question, 183–184
 - evaluating arguments, 178–179
 - questioning arguments, 176–178
- Condemnation of Socrates, 1–3
- Conscientious feelings of mankind, 162–165
- Consciousness
 - failure of Behaviorism, 73
 - Functionalism and, 79–80
 - idea of immortal soul and, 40
 - in machines. *See* Artificial intelligence (AI)
 - personhood and, 53
 - split brains and, 56–57
- Control
 - Determinist Argument, 109–110
 - happiness and, 169
- Creation
 - of evil, 27–29
 - free will and moral character, 33–37
 - God and evil, 29–30
 - of universe. *See* Origins of universe
- Creation science, 19–20
- Creativity, 85–86
- Credit-eliminating conditions, xi
- Crime
 - Social Contract Theory, 159
 - Utilitarianism and, 161
- Cultural differences, 140–144
- Custom, 140–144
- D**
- Darley, J. M., 103–104
- Darrow, Clarence, 94–96, 107
- Darwin, Charles
 - evolution and intelligent design, 15–20
 - on suffering of animals, 37
- Darwin's Black Box: The Biochemical Challenge to Evolution* (Behe), 19
- Death
 - identity and, 65–66
 - life after. *See* Afterlife
 - meaning of life, 169–171
 - Same-Brain Theory and, 60–61
 - of Socrates, 3–9
- Decision-making
 - argument that we cannot predict, 113–114
 - Determinism and, 99
- Deep Blue, 86
- Defeasible, 136
- Defense response to problem of evil, 29–30
- Defining God, 24–25
- Delgado, Jose
 - Argument from Experience, 111
 - Determinism, 98–99
- Deliberation and free will, 119–120
- Democracy, 2–3

- Demons, 126–127
- Descartes, René
 - Mind-Body Dualism, 67–71
 - Ontological Argument, 25
 - problem of reality, 126–127
 - Ryle's criticism of, 72
 - Theological response
 - to problem of reality, 130–133
 - on thinking and machines, 83
- Design, Argument from, 11–15
- Desire
 - Compatibilism, 116
 - problem with Compatibilism, 118
- Determinism
 - argument of, 109–110
 - defined, 96–99
 - ethics and free will, 118–124
 - Libertarian response, 110–115
- Dialogues Concerning Natural Religion* (Hume), 13–14
- Dialogues of Plato, 1
- Dilemma, false, 184
- Direct Realism, 133–134
- Disagreement, Argument from, 144–146
- Divine Command Theory, 154–157
- Dreams
 - Memory-Links Theory, 63
 - in near-death experiences, 43
 - reality and, 126–127, 132
- Drugs, 43
- Dualism, Mind-Body, 67–69
- Duplicator, 52–53
- Dybvig, Kevin, xii
- E**
- Edward, John, 49
- Edwards, Paul, 45–46
- Ego. *See* Personal identity
- Einstein, Albert, 112
- Eliza, 89–90
- Elliott, Heather, xi–xii
- emergence, radical, 70–71
- enjoyment, 173
- Enlightenment, 29
- An Enquiry Concerning Human Understanding* (Hume), 50
- Environment
 - in Determinism, 100
 - genes and behavior, 107–108
 - happiness and, 168
- Epicetetus, 167–168
- Epicurus
 - on death, 170
 - meaning of life, 167
- Epidermolysis bullosa, 30–31
- Equivocation, 180–181
- Ethics. *See also* Morality
 - appeal to authority, 182
 - free will and, 118–124
 - importance of human interests, 151–152
 - religion and, 154–157
 - science and, 144–151
 - as social convention, 140–144
 - Thrasymachus's challenge, 139–140
- Europe, 10
- Euthyphro* (Plato), 156
- Evaluating arguments, 176–184
- Evidence
 - for afterlife, 41–49
 - Argument from Lack of Proof, 146–148
 - of artificial intelligence, 93
 - problem of evil, 28

Same-Evidence Argument,
 12–13
 for views about knowledge,
 129–130
 Evil
 free will and moral character,
 33–37
 God and, 29–32
 why do good people suffer,
 27–29
 Evolution, 15–20
Excitable Cortex in Conscious Man,
 The (Penfield), 98
 Excuses, 121
 Existence of God. *See* God
 Experience
 argument against
 Determinism, 111
 Direct vs. Indirect Realism,
 133–134
 problems with materialist
 theories, 80–81
 Eyes
 evolution of, 19–20
 perception, 136–137
F
 Fallacies, 180–184
 False dilemma, 184
 Fatalism
 defined, 97
 free will and, 119
 Fear, 30
 First Cause Argument,
 21–23, 24
 Flawed arguments, 178–179
 Forgetting, 61–62
 Fox, Katie, 47
 Fox, Maggie, 47
 Free rider problem,
 158–159

Free will
 are people responsible,
 94–96
 begging the question,
 183–184
 Compatibilist response,
 115–118
 Determinism, 96–99
 Determinist Argument,
 109–110
 ethics and, 118–124
 evil and, 33–37
 genes and behavior,
 104–108
 Libertarian response,
 110–115
 psychology, 99–104
 Free Will Defense, 33–34
 Freudianism, 102
Fugitive, The, 8
 Functionalism
 defined, 78–79
 problem with, 79–80

G

Gallup poll, 10
 Gaunilo, 25
 Genetics
 behavior and, 104–108
 happiness and, 168
 Geometrical increase of
 populations, 16–18
 Geons, 136
 Giggie Sisters, 105
 Glaucon, 153–155
 God
 Argument from Design,
 11–15
 conclusion, 26
 ethics and, 140, 154–157
 evil and, 29–32

God (*continued*)
 evolution and intelligent design, 15–20
 Fatalism and, 97
 First Cause Argument, 21–23
 Free Will Defense, 33–37
 Idealism and, 129, 130
 indifferent universe and, 171–173
 is it reasonable to believe, 10–11
 as necessary being, 23–25
 Utilitarianism and, 161
 why do good people suffer, 27–29
 Goodness. *See also* Morality
 evaluating people's, 120
 evil as necessary to appreciate, 31
 Good Samaritan experiment, 103–104
 Government
 idea of social contract, 6–9
 why Socrates had to die, 3–9
 Gravity, 70
 Guards-and-prisoners experiment, 103
 Gun control, 145
 Gyges, Ring of, 153–154

H

Happiness
 meaning of life, 167–169
 problem of point of view, 166–167
 Utilitarianism, 160–162
 Harris, Judith Rich, 107
 Heaven, 154–155
 Hell, 154–155
 Heritability of traits, 16–18

Herodotus, 140
 Hobbes, Thomas
 free will, 115
 idea of social contract, 6
 Social Contract Theory, 157
 Hollingshead, Daniel, xii
 Homosexuality
 appeal to authority, 182
 ethics and religion, 156
 Utilitarianism and, 162
 Hot reading, 49
 Houdini, 48
 Humanity
 conscientious feelings of mankind, 162–165
 ethics and free will, 118
 ethics and objectivity, 151–152
 free will and moral character, 33–37
 Hume, David
 on altruism, 160
 argument against miracles, 49–51
 Bundle Theory, 54–55
 Darwin and, 20
 on ethics, 140
 First Cause Argument and, 22
 free will, 115
 objects to Argument from Design, 13–15
 on values, 149
 Huxley, T. H., 70

I

"I." *See* Personal identity
 Idealism, 127–129
 Identity. *See* Personal identity
 Ignorance
 responsibility and, 121
 Socrates's view of, 2

Illness, 121

Illusion

Direct vs. Indirect Realism,
133–134

meaning of life, 174

perception and, 125–127

Immaterial substances

of body and mind, 67

mind as, 70

Immortality. *See* Morality

Immortal soul, 38–41

Impartiality of Utilitarianism,
162–164

Inception, 126

Indifferent universe, 171–173

Indirect Realism, 133–134

Inference of God's existence,
14–15

Intelligence, artificial. *See*

Artificial intelligence (AI)

Intelligent design

argument for God, 11–15

Theory of Natural Selection
and, 18–20

Intensification of function, 20

Intentionality, 81–82

Investigator bias, 42

Islam

indifferent universe and,
171

Sharia, 141

J

Jackson, Frank, 80–81

James, Henry, 48

James, William, 48

Job

notes on, 186–187

suffering of, 27–29

Johnson, Samuel, 111

Judaism, 171

Justice. *See also* Ethics

relativism and, 140–144

Thrasymachus's challenge,
139–140

K

Kant, Immanuel

on ethics and religion,
156–157

free will, 115

Ontological Argument, 25

vision and the brain, 134

Kasparov, Garry, 86

Knowledge of world around us

conclusion, 137–138

Descartes' theological
response, 130–133

Direct vs. Indirect Realism,
133–134

evidence for views, 129–130

Idealism, 127–129

problem of perception,
125–127

vision and the brain, 134–137

Kornhuber, H. H., 99

L

Lack of Proof, Argument from,
146–148

Language, Truth and Logic
(Ayer), 44

Laplace, Pierre-Simon, 97–98

Lawal, Amina, 141–144

Laws

moral, 154–157

Social Contract Theory, 159
why Socrates had to die, 3–9

Laws of Nature, 110–111,
111–113

Learning, 173–174

Leibniz, Gottfried Wilhelm, 25

Leopold, Nathan, 94–96, 107

Libertarian response to
 Determinist Argument,
 110–115

Life after death. *See* Afterlife

Life after Life (Moody), 41–44

Life as prelude to eternal life,
 35–36

Life, meaning of. *See* Meaning
 of life

Light in near-death
 experiences, 41–44

Locke, John
 memory and responsibility, 63
 Memory Theory, 61
 on Same-Body Theory, 59
 on sense organs, 74

Loeb, Richard, 94–96, 107

Logic, 176–184. *See also* Reasoning

Logical problem of evil, 28

Love, 172

Lying, 151–152

M

Machine thought. *See* Artificial
 intelligence (AI)

Magic, 48

Mai, Mukhtar, 141–144

Marx, Groucho, 173

Materialist theories
 defined, 71
 doubts about, 80–82
 of mind, 71–80

Matrix, The, 126

Meaning of life
 death, 169–171
 happiness, 167–169
 particular, 173–175
 problem of point of view,
 166–167
 religion and indifferent
 universe, 171–173

Meditations on First Philosophy
 (Descartes), 69

Mediums, 47–49

Memory-Links Theory
 defined, 62–63
 triviality of, 65

Memory Theory
 defined, 61–65
 triviality of, 64–65

Mental activity
 Bundle Theory of
 personhood, 54–55
 decision-making, 98–99
 Libertarianism and free will,
 115
 of machines. *See* Artificial
 intelligence (AI)
 memory and personhood,
 61–65
 problem of mind and body.
 See Body and mind
 split brains, 55–57

Metaphysical Argument,
 148–151

Method, Socratic, 4

Mexico, 169

Milgram, Stanley, 103

Mill, John Stuart
 free will, 115
 on happiness, 169
 impartiality of Utilitarianism,
 163
 meaning of life, 166–167
 Utilitarianism, 160–161

Mind
 body and. *See* body and mind
 split brains, 55–57

Mind-Body Dualism
 defined, 67–69
 Elizabeth's objections to,
 69–70

- Mind–Brain Identity Theory
 - advantages of, 77–78
 - defined, 74–75
 - tokens and types, 76–77
- Minnesota Study of Twins
 - Reared Apart, 105–106
- Miracles, 49–51
- Mistakes, 121
- Money and happiness, 168–169
- Moody, Raymond A. Jr., 41–44
- Moralities, 150
- Morality
 - benevolence and, 159–164
 - conclusion, 164–165
 - equivocation of Socrates, 180–181
 - ethics. *See* Ethics
 - evaluating people's, 120
 - evil and character, 33–37
 - is it a matter of luck, 102–104
 - religion and, 154–157
 - responsibility and free will, 122–124
 - Ring of Gyges, 153–154
 - Social Contract Theory, 157–159
- Mortality
 - identity and, 66
 - meaning of life, 169–171
- Murder
 - begging the question, 183–184
 - ethics and, 145, 149
 - social contract against, 159
- Murphy, Bridey, 45–46
- Mystery of afterlife, 41
- N**
- Nagel, Thomas, 167
- Narveson, Jan, 163
- Natural evil, 34
- Natural Selection, 15–20
- Natural Theology* (Paley), 18
- Natural Theory, xi
- Natural wonders, 11–12
- Near-death experiences, 41–44
- Neisser, Ulric, 64
- Nelson, Kevin, 44
- New Testament*, 39–40
- Nietzsche, Friedrich, 149–150
- Nigeria, 141–142
- Non-human animals
 - suffering of, 37
 - values and, 149
- Numerical identity
 - cell regeneration and, 60
 - defined, 58
 - memory and responsibility, 64
- O**
- Objectivity and ethics
 - importance of human interests, 151–152
 - science and, 144–151
 - social convention, 140–144
 - Thrasymachus's challenge, 139–140
 - Utilitarianism, 163–164
- Objects. *See* Physical objects
- On the Origin of Species* (Darwin), 16
- Ontological Argument
 - defined, 25
 - Descartes' conclusions and, 132
- Origins of universe
 - Argument from Design, 11–15
 - evolution and intelligent design, 15–20
 - First Cause Argument, 21–23
 - necessity of God, 23–25
- Oxygen deprivation, 43

P**Pain**

- in brain, 74–75
- Functionalism and, 78–79
- as necessary, 30–31
- tokens and types, 76–77

Pakistan, 141–142**Paley, William**

- Argument from Design, 12–13
- Theory of Natural Selection and, 18

Parfit, Derek

- on identity, 66
- on split brains, 56–57

Penfield, Wilder

- Determinism, 98
- Mind–Brain Identity Theory, 74–75

Perception

- conclusion, 137–138
- Direct vs. Indirect Realism, 133–134
- problem of reality, 125–127
- vision and the brain, 134–137

Perfection, 34–35**Personal identity**

- bodily continuity, 58–61
- conclusion, 65–66
- memory, 61–65
- Mind–Brain Identity Theory, 74–75
- personhood at a time, 53–57
- personhood over time, 57–58
- problem of, 52–53

Personality, 104–108**Personal relationships, 173****Personhood**

- at a time, 53–57
- over time, 57–58

Philosophy

- overview of, x
- religion and, 29

Physical body

- continuity and personhood, 58–61
- problem of mind and body. *See* Body and mind
- warning system, 30–31

Physical objects

- Direct vs. Indirect Realism, 133–134
- existence of, 127–129
- perceiving with brain, 134–137

Physical pain. *See* Pain**Piecemeal-Replacement****Argument**

- defined, 84–85
- objections, 85–87

Pinker, Steven, 135**Plantinga, Alvin, 29****Plato**

- Ring of Gyges, 153–154
- Socrates and, 1
- Symposium*, 9
- Thrasymachus's challenge, 139

Plutarch, 59–60**Point of view, 166–167****Polemarchus, 180–181****Politics, 2****Praise**

- morality and free will, 124
- responsibility and free will, 121

Predictability

- Compatibilism and, 116–117
- Determinism and, 113–114

Premise

- begging the question, 183–184
- defined, 176–178

Pre-Socratic, 3
 Prince and the cobbler, 59
 Princess Elizabeth of Bohemia, 69–70
 Principle of Utility, 160–162
 Privacy of mental states, 68
 Problem of evil. *See* Evil
 Problem of personal identity. *See* Personal identity
 Problem of point of view, 166–167
 Problem of radical emergence, 70–71
 Programming computers, 85–86
 Proof. *See* Evidence
 Psychic communication with the dead, 47–49
 Psychology
 Behaviorism, 71–74
 case against free will, 99–104
 doubts about materialist theories, 80–82
 Functionalism, 78–80
 genes and behavior, 104–108
 immortal soul and, 40
 of memory, 64–65
 of near-death experiences, 42
 responsibility and free will, 94–96
Psychology from an Empirical Standpoint (Brentano), 81
 Public physical states, 68
 Punishment, 32

Q

Qualia, 127
 Qualitative identity
 defined, 58
 memory and responsibility, 64
 Quantum mechanics, 112

Questioning arguments
 begging the question, 183–184
 reasoning, 176–178
 Quine, W. V., x

R

Rachels, Carol, xii
 Radical emergence, 70–71
 Randi, James, x, 48
 Randomness, 112
 Ransom, Champe, 46–47
 Rape, 141–144
 “Rapid eye movement” (REM) sleep, 43–44
 Ready, W. B., 45
 Realism, Direct vs. Indirect, 133–134
 Reality
 conclusion, 137–138
 Descartes’ theological response, 130–133
 Direct vs. Indirect Realism, 133–134
 evidence for views, 129–130
 Idealism, 127–129
 moral, 148–151
 problem of illusion, 125–127
 vision and the brain, 134–137
 Reasoning
 ethics and religion, 156
 evaluating arguments, 176–184
 meaning of life, 174–175
 moral, 150–151
 overview of, x
 Reid, Thomas, 61–62
 Reincarnation
 evidence for, 45–47
 Hume’s argument against miracles, 51

Relativism in ethics, 140–144

Religion

appeal to authority, 182

belief in God. *See* God

Descartes' conclusions about
reality, 130–133

God and evil, 29–32

idea of immortal soul,
38–40

indifferent universe and,
171–173

morality and, 154–157

REM (“rapid eye movement”)
sleep, 43–44

Republic (Plato)

equivocation in, 180–181

Ring of Gyges, 153–154

Thrasymachus's challenge,
139

Responsibility

case against free will, 94–96

Determinist Argument,
109–110

ethics and free will, 120–124

free will and moral character,
33–37

memory and, 63–64

Right and wrong. *See* Ethics

Ring of Gyges, 153–154

Robots, 119. *See also* Artificial
intelligence (AI)

Rousseau, Jean-Jacques, 6

Rudolph, Eric

Determinism, 101

evaluating people's morality,
120

Rules, moral. *See* Morality

Russell, Bertrand

on Direct Realism, 134

on responsibility, 120–121

Ryle, Gilbert, 72

S

Sacks, Oliver, 136

Same-Body Theory, 58–60

Same-Brain Theory, 60–61

Same-Evidence Argument
defined, 12–13

Hume's objections to, 14

Science

argument against

Determinism, 111–113

Behaviorism, 71–73

Determinism, 97–99

ethics and, 140, 144–151

explaining near-death
experiences, 43–44

Hume's argument against
miracles, 49–51

origins of universe. *See*
Origins of universe

split brains, 55–57

stories of reincarnation,
46–47

Search for Bridey Murphy, The
(Bernstein), 45

Searle, John, 91

Self. *See* Personal identity

Self-interest

ethics and religion, 154–157

morality and benevolence,
159–164

Ring of Gyges, 153–154

Social Contract Theory,
157–159

Semantics, 92–93

Sense-datum, 127

Senses, 74. *See also* Perception

Sentencing of Socrates, 3

Sharia, 141

Sherzer, Alex, xii

Ship of Theseus, 59–60

Sight, 134–137

- Sin, 32
- Singer, Isaac Bashevis, 109
- Singer, Peter, 164
- Skinner, B. F.
- Behaviorism, 71–72
 - Determinism and, 101–102
- Slave morality, 150
- Social conditions in
- Determinism, 100
- Social Contract Theory
- impartiality of, 162–163
 - morality and, 157–159
 - Socrates's argument, 6–9
- Social convention of ethics, 140–144
- Socrates
- condemnation, 1–3
 - death and legacy, 3–9
 - equivocation of, 180–181
 - on ethics and religion, 155–156
 - flawed argument, 179–180
 - idea of immortal soul, 38–40
 - meaning of life, 174
 - questioning arguments, 176–178
 - Thrasymachus's challenge, 139–140
- Sophists, 139
- Soul
- idea of immortal, 38–40
 - mind-body dualism, 67–69
- Soul Theory
- vs. Bundle Theory, 53–55
 - conclusion, 65–66
- Soundness of arguments, 177
- Sperry, Roger W., 55–56
- Spiritualism, 47–49
- St. Anselm, 24
- St. Augustine
- evil as necessary, 31
 - Fatalism, 97
- St. Paul, 40
- State
- analogy between one's parents and, 5–6
 - argument about destroying, 4–5
- Statutes of limitation, 64
- Stevenson, Ian, 46–47
- Stöhr, Oskar, 105–106
- Stoics, 167–168
- Struggle for existence, 17
- Subjection of Women, The* (Mill), 161
- Subjectivity
- of ethics, 152
 - problems with materialist theories, 80–81
- Substances
- of body and mind, 67
 - problem of radical emergence, 70–71
- Suffering
- Free Will Defense, 33–37
 - as necessary, 30–31
 - as part of development of moral character, 34–35
 - as punishment, 32
 - why good people experience, 27–29
- Survival value, 18
- Surviving death. *See* Afterlife
- Symposium* (Plato), 9
- Syntax, 92–93
- T**
- Technology and intelligence.
- See* Artificial intelligence (AI)

Thales, 3
 Theodicy, 29–30
 Theological response to
 problem of perception,
 130–133
 Theories
 Bundle. *See* Bundle Theory
 Divine Command Theory,
 154–157
 impartiality of Social
 Contract Theory, 162–163
 Memory-Links Theory,
 62–63, 65
 Mind-Brain Identity Theory.
 See Mind-Brain Identity
 Theory
 Natural Theory, xi
 Same-Body Theory, 58–60
 Same-Brain Theory, 60–61
 Social Contract Theory,
 157–159
 Soul Theory, 53–55, 65–66
 Theory of Natural Selection,
 15–20
 Theron, Charlize, 182
 Thinking. *See* Mental activity
Thirteenth Floor, The, 126
 Thirty Tyrants, 2–3
 Thrasymachus's challenge,
 139–140
 Tighe, Virginia, 45–46
 Time
 personhood at a, 53–57
 personhood over, 57–58
 Tipping points, 86–87
 Tokens, 76–77
 Tolerance, 141
Total Recall, 126
 Transitive relations, 62
 Trial of Socrates, 1–3
 Trust, 152

Truth, objective ethical. *See*
 Objectivity and ethics
 Tunnel in near-death
 experiences, 41–44
 Turing, Alan, 87–91
 Turing Test
 can machines think, 87–91
 why it fails, 91–93
 Twin studies, 105–106
 Types, 76–77

U

United Nations Children's Fund
 (UNICEF), 162–163
 United States
 beliefs, 41
 ethics of, 142
 happiness in, 169
 religion in, 10
 Universe
 Determinism and, 111–113
 origins. *See* Origins of
 universe
 religion and indifferent,
 171–173
 Utilitarianism
 defined, 160–162
 impartiality of, 162–164

V

Validity of arguments, 177
 Values
 cultural differences,
 140–144
 meaning of life, 170–171
 meaning of particular lives,
 173–175
 Metaphysical Argument,
 148–151
 religion and indifferent
 universe, 172

Van Inwagen, Peter
 Compatibilism and, 118
 on eternal life, 35–36
 necessity of God, 23
 Variation
 in evolution of eye, 19–20
 Theory of Natural Selection,
 16–18
 Vats, 125–126
 Vice, 149
 Violence and biology, 106–108
 Virtual reality, 125–127
 Virtue. *See also* Morality
 evil and development of,
 34–35
 meaning of life, 174
 Vision and the brain, 134–137

W

Walton, Mark, xii
 Watson, John B., 71

Watson, Thomas J., 83–84
 Wealth and happiness,
 168–169
 Weizenbaum, Joseph, 90
 West Africa, 10
*What to Do When You Are Dead:
 Living Better in the Afterlife*
 (Hamilton-Parker), 42
 Wittgenstein, Ludwig, 134
 Wonders of nature, 11–12
 Wrenn, Chase, xii
 Wrong and right. *See* Ethics
 Wrongdoing and punishment,
 32

Y

Yeats, W. B., 171
 Yufe, Jack, 105–106

Z

Zimbardo, Philip, 103

