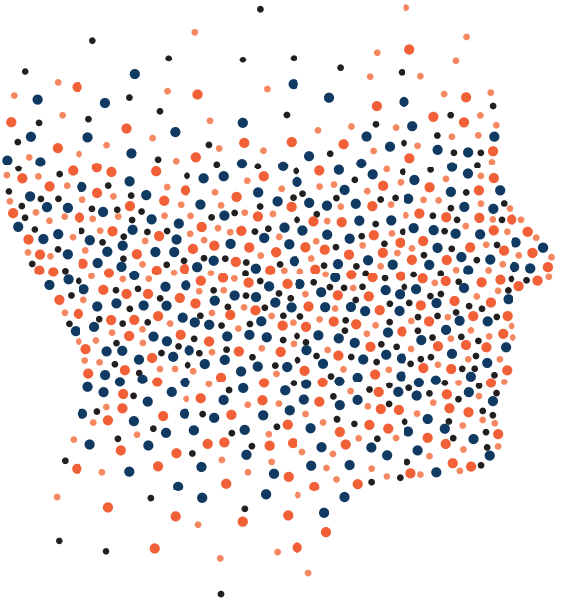
From the bestselling author of *Behave*

ROBERT SAPOLSKY

'One of the best scientist-writers of our time'
OLIVER SACKS



Determined

The Science of Life Without Free Will

'Moving, absorbing, compassionate'
OLIVER BURKEMAN, *OBSERVER*

‘One of the finest natural-history writers around’
New York Times

‘Fascinating, provocative and profound. This book tackles all sorts of big issues: how the human brain works, what makes us different, and what underlies everything we do. If Sapolsky is right, we might need to rethink justice and law, and for each of us personally, what it really takes to be happy and successful’
Daniel M. Davis, author of *The Secret Body*

‘Sapolsky’s decades of experience studying the effects of the interplay of genes and the environment on behaviour shine brightly . . . He provides compelling examples that bad luck compounds . . . [and] convincingly argues against claims that chaos theory, emergent phenomena or the indeterminism offered by quantum mechanics provide the gap required for free will to exist’
Science

‘Fascinating and challenging – though I’m not sure if I really had a say in the matter’
New Scientist

‘The information that he delivers us is tremendously valuable, interesting and in many cases actionable as well’
Andrew Huberman, *Huberman Lab* podcast

‘A complete legend’
Chris Williamson, *Modern Wisdom* podcast

‘One of the most revered scientists alive today’
Timothy Revell, *CultureLab* podcast

‘One of my intellectual heroes’
Scott Barry Kaufman, *The Psychology Podcast*

Copyrighted Material

Copyrighted Material

ROBERT M. SAPOLSKY

Robert M. Sapolsky is a professor of biology and neurology at Stanford University and the recipient of a MacArthur Foundation 'Genius Grant'. His previous books include the international bestseller *Behave: The Biology of Humans at Our Best and Worst*, *The Trouble with Testosterone*, *Why Zebras Don't Get Ulcers* and *A Primate's Memoir*.

Copyrighted Material

ALSO BY ROBERT M. SAPOLSKY

Behave: The Biology of Humans at Our Best and Worst

Monkeyluv: And Other Essays on Our Lives as Animals

A Primate's Memoir

The Trouble with Testosterone and Other Essays on the Biology of the Human Predicament

Why Zebras Don't Get Ulcers: A Guide to Stress, Stress-Related Diseases, and Coping

Stress, the Aging Brain, and the Mechanisms of Neuron Death

Copyrighted Material

ROBERT M. SAPOLSKY

Determined

The Science of Life Without Free Will

Copyrighted Material

VINTAGE

1 3 5 7 9 10 8 6 4 2

Vintage is part of the Penguin Random House group of companies whose addresses can be found at global.penguinrandomhouse.com

Penguin Random House UK

First published in Vintage in 2023
First published in Great Britain by The Bodley Head in 2023

Copyright © Robert M. Sapolsky

Robert M. Sapolsky has asserted his right to be identified as the author of this Work in accordance with the Copyright, Designs and Patents Act 1988

The English translation by Daniel Kahn, of the Yiddish poem 'Mayn Rue Platz' by Morris Rosenfeld, on p. 402 is used by permission.
Pages 491–92 constitute an extension of this copyright page.

penguin.co.uk/vintage

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

The authorised representative in the EEA is Penguin Random House Ireland, Morrison Chambers, 32 Nassau Street, Dublin D02 YH68

A CIP catalogue record for this book is available from the British Library

ISBN 9781529920062

Penguin Random House is committed to a sustainable future for our business, our readers and our planet. This book is made from Forest Stewardship Council® certified paper.

MIX
Paper | Supporting responsible forestry
FSC® C018179
www.fsc.org

Copyrighted Material

To L, and to B & R,

Who make it all seem worth it.

Who make it worth it.

Copyrighted Material

Copyrighted Material

CONTENTS

1. Turtles All the Way Down *1*

2. The Final Three Minutes of a Movie *18*

3. Where Does Intent Come From? *45*

4. Willing Willpower: The Myth of Grit *85*

5. A Primer on Chaos *125*

6. Is Your Free Will Chaotic? *144*

7. A Primer on Emergent Complexity *154*

8. Does Your Free Will Just Emerge? *192*

9. A Primer on Quantum Indeterminacy *203*

10. Is Your Free Will Random? *214*

10.5. Interlude *240*

Copyrighted Material

11. Will We Run Amok? *245*

12. The Ancient Gears within Us:
How Does Change Happen? *268*

13. We Really Have Done This Before *300*

14. The Joy of Punishment *341*

15. If You Die Poor *385*

Acknowledgments *405*

Appendix: Neuroscience 101 *407*

Notes *431*

Illustration Credits *491*

Index *493*

Copyrighted Material

Determined

Copyrighted Material

my brain: click them

me: why?

my brain: you gotta



Copyrighted Material

1

Turtles All the Way Down

When I was in college, my friends and I had an anecdote that we retold frequently; it went like this (and our retelling was so ritualistic that I suspect this is close to verbatim, forty-five years later):

> So, it seems that William James was giving a lecture about the nature of life and the universe. Afterward, an old woman came up and said, "Professor James, you have it all wrong."
>
> To which James asked, "How so, madam?"
>
> "Things aren't at all like you said," she replied. "The world is on the back of a gigantic turtle."
>
> "Hmm." said James, bemused. "That may be so, but where does that turtle stand?"
>
> "On the back of another turtle," she answered.
>
> "But madam," said James indulgently, "where does that turtle stand?"
>
> To which the old woman responded triumphantly: "It's no use, Professor James. It's turtles all the way down!"*

*The "turtles all the way down" story has versions featuring other celebrated thinkers as the fall guy, rather than William James. We told our version because we liked James's beard, and there was a building on campus named for him. The "turtles all the way down" punch line has been referenced in numerous cultural contexts, including a great book with that title by John Green

Copyrighted Material

Oh, how we loved that story, always told it with the same intonation. We thought it made us seem droll and pithy and attractive.

We used the anecdote as mockery, a pejorative critique of someone clinging unshakably to illogic. We'd be in the dinner hall, and someone had said something nonsensical, where their response to being challenged had made things worse. Inevitably, one of us would smugly say, "It's no use, Professor James!" to which the person, who had heard our stupid anecdote repeatedly, would inevitably respond, "Screw you, just listen. This actually makes sense."

Here is the point of this book: While it may seem ridiculous and nonsensical to explain something by resorting to an infinity of turtles all the way down, *it actually is much more ridiculous and nonsensical to believe that somewhere down there, there's a turtle floating in the air.* The science of human behavior shows that turtles can't float; instead, it is indeed turtles all the way down.

Someone behaves in a particular way. Maybe it's wonderful and inspiring, maybe it's appalling, maybe it's in the eye of the beholder, or maybe just trivial. And we frequently ask the same basic question: Why did that behavior occur?

If you believe that turtles can float in the air, the answer is that it just happened, that there was no cause besides that person having simply decided to create that behavior. Science has recently provided a much more accurate answer, and when I say "recently," I mean in the last few centuries. The answer is that the behavior happened because something that preceded it caused it to happen. And why did that prior circumstance occur? Because something that preceded it caused *it* to happen. It's antecedent causes all the way down, not a floating turtle or causeless cause to be found. Or as Maria sings in *The Sound of Music,* "Nothing comes from nothing, nothing ever could."*

(Dutton Books, 2017). All the versions of the story have a male Philosopher King Whoever being challenged by an absurd old woman, which now seems kind of sexist and ageist. That didn't particularly register with us then, we adolescent males of our time and place.

Copyrighted Material

*My wife is a musical theater director, and I'm her rusty rehearsal pianist/generalized gofer; as a result, this book is riddled with allusions to musicals. If my college self, being ostensibly cool by

To reiterate, when you behave in a particular way, which is to say when your brain has generated a particular behavior, it is because of the determinism that came just before, which was caused by the determinism just before that, and before that, all the way down. The approach of this book is to show how that determinism works, to explore how the biology over which you had no control, interacting with environment over which you had no control, made you you. And when people claim that there are causeless causes of your behavior that they call "free will," they have (a) failed to recognize or not learned about the determinism lurking beneath the surface and/or (b) erroneously concluded that the rarefied aspects of the universe that do work indeterministically can explain your character, morals, and behavior.

Once you work with the notion that every aspect of behavior has deterministic, prior causes, you observe a behavior and can answer why it occurred: as just noted, because of the action of neurons in this or that part of your brain in the preceding second.[†] And in the seconds to minutes before, those neurons were activated by a thought, a memory, an emotion, or sensory stimuli. And in the hours to days before that behavior occurred, the hormones in your circulation shaped those thoughts, memories, and emotions and altered how sensitive your brain was to particular environmental stimuli. And in the preceding months to years, experience and environment changed how those neurons function, causing some to sprout new connections and become more excitable, and causing the opposite in others.

And from there, we hurtle back decades in identifying antecedent

---

referring to William James, had been told that my future included my family and me debating who was the greatest Elphaba of all time,[*] I would have been astonished—"Musicals? Broadway MUSICALS?! What about atonalism?" It's not what I asked for; sometimes life just slips in through a back door.

([*]Idina Menzel. Obviously.)

[†]The appendix is an introduction to neuroscience, for readers without a background in this area. Also, anyone who has read an agonizingly long book that I wrote (*Behave: The Biology of Humans at Our Best and Worst*, Penguin Press, 2017) will recognize the book summarized in the next few paragraphs: Why did that behavior occur? Because of events one second before, one minute . . . one century . . . one hundred million years before.

Copyrighted Material

causes. Explaining why that behavior occurred requires recognizing how during your adolescence a key brain region was still being constructed, shaped by socialization and acculturation. Further back, there's childhood experience shaping the construction of your brain, with the same then applying to your fetal environment. Moving further back, we have to factor in the genes you inherited and their effects on behavior.

But we're not done yet. That's because everything in your childhood, starting with how you were mothered within minutes of birth, was influenced by culture, which means as well by the centuries of ecological factors that influenced what kind of culture your ancestors invented, and by the evolutionary pressures that molded the species you belong to. Why did that behavior occur? Because of biological and environmental interactions, all the way down.*

As a central point of this book, those are all variables that you had little or no control over. You cannot decide all the sensory stimuli in your environment, your hormone levels this morning, whether something traumatic happened to you in the past, the socioeconomic status of your parents, your fetal environment, your genes, whether your ancestors were farmers or herders. Let me state this most broadly, probably at this point too broadly for most readers: we are nothing more or less than the cumulative biological and environmental luck, over which we had no control, that has brought us to any moment. You're going to be able to recite this sentence in your irritated sleep by the time we're done.

There are all sorts of aspects about behavior that, while true, are not relevant to where we're heading. For example, the fact that some criminal behavior can be due to psychiatric or neurological problems. That some kids have "learning differences" because of the way their brains work. That some people have trouble with self-restraint, because they grew up without any decent role models or because they're still a teenager with a

*"Interactions" implies that those biological influences are meaningless outside the context of social environment (as well as the reverse). They're inseparable. My orientation happens to be biological, and analyzing the inseparability from that angle is clearest in my mind. But at times, framing the inseparability from a biological rather than a social science perspective makes things clunky; I've tried to avoid that to the best of my biologist abilities.

Copyrighted Material

teenager's brain. That someone has said something hurtful merely because they're tired and stressed, or even because of a medication they're taking.

All of these are circumstances where we recognize that sometimes, biology can *impinge on* our behavior. This is essentially a nice humane agenda that endorses society's general views about agency and personal responsibility but reminds you to make exceptions for edge cases: judges should consider mitigating factors in criminals' upbringing during sentencing; juvenile murderers shouldn't be executed; the teacher handing out gold stars to the kids who are soaring in learning to read should do something special too for that kid with dyslexia; college admissions officers should consider more than just SAT cutoffs for applicants who have overcome unique challenges.

These are good, sensible ideas that should be instituted if you decide that *some* people have much less self-control and capacity to freely choose their actions than average, and that at times, we *all* have much less than we imagine.

We can all agree on that; however, we're heading into very different terrain, one that I suspect most readers will not agree with, which is deciding that we have *no* free will at all. Here would be some of the logical implications of that being the case: That there can be no such thing as blame, and that punishment as retribution is indefensible—sure, keep dangerous people from damaging others, but do so as straightforwardly and nonjudgmentally as keeping a car with faulty brakes off the road. That it can be okay to praise someone or express gratitude toward them as an instrumental intervention, to make it likely that they will repeat that behavior in the future, or as an inspiration to others, but never because they *deserve* it. And that this applies to you when you've been smart or self-disciplined or kind. Oh, as long as we're at it, that you recognize that the experience of love is made of the same building blocks that constitute wildebeests or asteroids. That no one has *earned* or is *entitled* to being treated better or worse than anyone else. And that it makes as little sense to hate someone as to hate a tornado because it supposedly decided to level your house, or

Copyrighted Material

to love a lilac because it supposedly decided to make a wonderful fragrance.

That's what it means to conclude that there is no free will. This is what I've concluded, for a long, long time. And even I think that taking that seriously sounds absolutely nutty.

Moreover, most people agree that it sounds that way. People's beliefs and values, their behavior, their answers to survey questions, their actions as study subjects in the nascent field of "experimental philosophy," show that people believe in free will when it matters—philosophers (about 90 percent), lawyers, judges, jurors, educators, parents, and candlestick makers. As well as scientists, even biologists, even many neurobiologists, when push comes to shove. Work by psychologists Alison Gopnik at UC Berkeley and Tamar Kushnir at Cornell shows that preschool kids already have a robust belief in a recognizable version of free will. And such a belief is widespread (but not universal) among a wide variety of cultures. We are not machines in most people's view; as a clear demonstration, when a driver or an automated car makes the same mistake, the former is blamed more.[1] And we are not alone in our faith in free will—research that we'll look at in a later chapter suggests that other primates even believe that there is free will.[2]

This book has two goals. The first is to convince you that there is no free will,* or at least that there is much *less* free will than generally assumed when it really matters. To accomplish that, we'll look at the way smart, nuanced thinkers argue *for* free will, from the perspectives of philosophy, legal thought, psychology, and neuroscience. I'll be trying to present their views to the best of my ability, and to then explain why I think they are all mistaken. Some of these mistakes arise from the myopia

*Some of the most extreme "there's NO free will" fellow travelers include philosophers such as Gregg Caruso, Derk Pereboom, Neil Levy, and Galen Strawson; I'll often be discussing their thinking in the pages to come; as an important point, while all reject free will in the everyday sense we understand it when justifying punishment and reward, their rejection is not particularly along biological grounds. In terms of rejecting free will almost entirely on biological grounds, my views are closest to those of Sam Harris, who, appropriately, is not only a philosopher, but a neuroscientist as well.

Copyrighted Material

(used in a descriptive rather than judgmental sense) of focusing solely on just one sliver of the biology of behavior. Sometimes this is because of faulty logic, such as concluding that if it's not possible to ever tell what caused X, maybe nothing caused it. Sometimes the mistakes reflect unawareness or misinterpretation of the science underlying behavior. Most interestingly, I sense that mistakes arise for emotional reasons that reflect that there being no free will is pretty damn unsettling; we'll consider this at the end of the book. So one of my two goals is to explain why I think all these folks are wrong, and how life would improve if people stopped thinking like them.[3]

Right around here, one might ask of me, Where do you get off? As will be seen, free-will debates often revolve around narrow issues—"Does a particular hormone actually cause a behavior or just make it more likely?" or "Is there a difference between wanting to do something and wanting to want something?"—that are usually debated by specialized authorities. My intellectual makeup happens to be that of a generalist. I'm a "neurobiologist" with a lab that does things like manipulate genes in a rat's brain to change behavior. At the same time, I spent part of each year for more than three decades studying the social behavior and physiology of wild baboons in a national park in Kenya. Some of my research turned out to be relevant to understanding how adult brains are influenced by the stress of childhood poverty, and as a result, I've wound up spending time around the likes of sociologists; another facet of my work has been relevant to mood disorders, leading me to hang with psychiatrists. And for the last decade, I've had a hobby of working with public defender offices on murder trials, teaching juries about the brain. As a result, I've been carpetbagging in a number of different fields related to behavior. Which I think has made me particularly prone toward deciding that free will doesn't exist.

Why? Crucially, if you focus on any single field like these—neuroscience, endocrinology, behavioral economics, genetics, criminology, ecology, child development, or evolutionary biology—you are left with plenty of wiggle room for deciding that biology and free will can coexist. In the words of UC San Diego philosopher Manuel Vargas, "Claiming that some scientific

Copyrighted Material

result shows the falsity of 'free will' . . . is either bad scholarship or academic hucksterism."[4] He is right, if in-your-face. As we will see in the next chapter, most experimental neurobiology research about free will is narrowly anchored by the result of one study that examined events that happen in the brain a few seconds before a behavior occurs. And Vargas would correctly conclude that this "scientific result" (plus the spin-offs it has generated in the subsequent forty years) doesn't prove there's no free will. Similarly, you can't disprove free will with a "scientific result" from genetics—genes in general are not about inevitability but, rather, about vulnerability and potential, and no single gene, gene variant, or gene mutation has ever been identified that falsifies free will;* you can't even do it when considering *all* our genes at once. And you can't disprove free will from a developmental/sociological perspective by emphasizing the scientific result that a childhood filled with abuse, deprivation, neglect, and trauma astronomically increases the odds of producing a deeply damaged and damaging adult—because there are exceptions. Yeah, no single result or scientific discipline can do that. But—and this is the incredibly important point—*put all the scientific result*s *together*, *from all the relevant scientific disciplin*es, and there's no room for free will.†

Why is that? Something deeper than the idea that if you examine enough different disciplines, one -ology after another, you're bound to

*That said, there are a few rare diseases that are guaranteed to alter behavior because of a mutation in a single gene (e.g., Tay-Sachs, Huntington's, and Gaucher diseases). Nonetheless, this isn't remotely related to issues of our everyday sense of free will, as these diseases cause massive damage in the brain.

†I'd like to note something in preparation for my spending the first half of the book repeatedly saying, "They're all wrong," about a lot of scholars thinking about this subject. I can be intensely emotional about ideas, with some evoking the closest I can ever feel to religious awe and others seeming so appallingly wrong that I can be bristly, acerbic, arrogantly judgmental, hostile, and unfair in how I critique them. But despite that, I am majorly averse to interpersonal conflict. In other words, with a few exceptions that will be clear, none of my criticisms are meant to be personal. And as a "some of my best friends" cliché, I like being around people with a particular type of belief in free will, because they're generally nicer people than those on "my side" and because I hope some of their peace will rub off on me. What I'm trying to say is that I hope I won't be sounding like a jerk at times, because I very much don't want to.

Copyrighted Material

eventually find one that provides a slam dunk, falsifying free will all by itself. It is also deeper than the idea that even though each discipline has a hole that precludes it from falsifying free will, at least one of the other disciplines compensates for it.

Crucially, all these disciplines collectively negate free will because they are all interlinked, constituting the same ultimate body of knowledge. If you talk about the effects of neurotransmitters on behavior, you are also implicitly talking about the genes that specify the construction of those chemical messengers, and the evolution of those genes—the fields of "neurochemistry," "genetics," and "evolutionary biology" can't be separated. If you examine how events in fetal life influence adult behavior, you are also automatically considering things like lifelong changes in patterns of hormone secretion or in gene regulation. If you discuss the effects of mothering style on a kid's eventual adult behavior, by definition you are also automatically discussing the nature of the culture that the mother passes on through her actions. There's not a single crack of daylight to shoehorn in free will.

As such, the first half of the book's point is to rely on this biological framework in rejecting free will. Which brings us to the second half of the book. As noted, I haven't believed in free will since adolescence, and it's been a moral imperative for me to view humans without judgment or the belief that anyone deserves anything special, to live without a capacity for hatred or entitlement. And I just can't do it. Sure, sometimes I can sort of get there, but it is rare that my immediate response to events aligns with what I think is the only acceptable way to understand human behavior; instead, I usually fail dismally.

As I said, even I think it's crazy to take seriously all the implications of there being no free will. And despite that, the goal of the second half of the book is to do precisely that, both individually and societally. Some chapters consider scientific insights about how we might go about dispensing with free-will belief. Others examine how some of the implications of rejecting free will are not disastrous, despite initially seeming that way.

Copyrighted Material

Some review historical circumstances that demonstrate something crucial about the radical changes we'd need to make in our thinking and feeling: *we've done it before.*

The book's intentionally ambiguous title reflects these two halves—it is both about the science of why there is no free will and the science of how we might best live once we accept that.

STYLES OF VIEWS: WHOM I WILL BE DISAGREEING WITH

I'm going to be discussing some of the common attitudes held by people writing about free will. These come in four basic flavors:*

The world is deterministic and there's no free will. In this view, if the former is the case, the latter has to be as well; determinism and free will are not compatible. I am coming from this perspective of "hard incompatibilism."†

The world is deterministic and there is free will. These folks are emphatic that the world is made of stuff like atoms, and life, in the elegant words of psychologist Roy Baumeister (currently at the University of Queensland in

*Note: I won't be considering any theologically based Judeo-Christian views about these subjects beyond this broad summary here. As far as I can tell, most of the theological discussions center around omniscience—if God's all-knowingness includes knowing the future, how can we ever freely, willingly choose between two options (let alone be judged for our choice)? Amid the numerous takes on this, one answer is that God is outside of time, such that past, present, and future are meaningless concepts (implying, among other things, that God could never relax by going to a movie and being pleasantly surprised by a plot turn—He always knows that the butler didn't do it). Another answer is one of the limited God, something explored by Aquinas—God cannot sin, cannot make a boulder too heavy for Him to lift, cannot make a square circle (or, as another example that I've seen offered by a surprising number of male but not female theologians, even God cannot make a married bachelor). In other words, God cannot do *everything*, He can just do whatever is possible, and foreseeing whether someone will choose good or evil is not knowable, even for Him. Related to this all, Sam Harris mordantly notes that even if we each have a soul, we sure didn't get to pick it.

†Which I'm viewing as synonymous with "hard determinism"; all sorts of philosophers, however, make fine distinctions between the two.

Copyrighted Material

Australia), "is based on the immutability and relentlessness of the laws of nature."[5] No magic or fairy dust involved, no substance dualism, the view where brain and mind are separate entities.* Instead, this deterministic world is viewed as compatible with free will. This is roughly 90 percent of philosophers and legal scholars, and the book will most often be taking on these "compatibilists."

The world is not deterministic; there's no free will. This is an oddball view that everything important in the world runs on randomness, a supposed basis of free will. We'll get to this in chapters 9 and 10.

The world is not deterministic; there is free will. These are folks who believe, like I do, that a deterministic world is not compatible with free will—however, no problem, the world isn't deterministic in their view, opening a door for free-will belief. These "libertarian incompatibilists" are a rarity, and I'll only occasionally touch on their views.

There's a related quartet of views concerning the relationship between free will and moral responsibility. The last word obviously carries a lot of baggage with it, and the sense in which it is used by people debating free will typically calls forth the concept of *basic desert,* where someone can *deserve* to be treated in a particular way, where the world is a morally acceptable place in its recognition that one person can deserve a particular reward, another a particular punishment. As such, these views are:

There's no free will, and thus holding people morally responsible for their actions is wrong. Where I sit. (And as will be covered in chapter 14, this is completely separate from forward-looking issues of punishment for deterrent value.)

There's no free will, but it is okay to hold people morally responsible for their

*Compatibilists make that clear. For example, one paper in the field is entitled "Free Will and Substance Dualism: The Real Scientific Threat to Free Will?" For the author, there's actually no threat to free will; there's a threat, though, of irksome scientists thinking they've scored points against compatibilists by labeling them as substance dualists. Because, to paraphrase a number of compatibilist philosophers, saying that free will doesn't exist because substance dualism is mythical is like saying that love doesn't exist because Cupid is mythical.

Copyrighted Material

actions. This is another type of compatibilism—an absence of free will and moral responsibility coexist without invoking the supernatural.

There's free will, and people should be held morally responsible. This is probably the most common stance out there.

There's free will, but moral responsibility isn't justified. This is a minority view; typically, when you look closely, the supposed free will exists in a very narrow sense and is certainly not worth executing people about.

Obviously, imposing these classifications on determinism, free will, and moral responsibility is wildly simplified. A key simplification is pretending that most people have clean "yes" or "no" answers as to whether these states exist; the absence of clear dichotomies leads to frothy philosophical concepts like partial free will, situational free will, free will in only a subset of us, free will only when it matters or only when it doesn't. This raises the question of whether the edifice of free-will belief is crumbled by one flagrant, highly consequential exception and, conversely, whether free-will skepticism collapses when the opposite occurs. Focusing on gradations between yes and no is important, since interesting things in the biology of behavior are often on continua. As such, my fairly absolutist stance on these issues puts me way out in left field. Again, my goal isn't to convince you that there's no free will; it will suffice if you merely conclude that there's so much less free will than you thought that you have to change your thinking about some truly important things.

Despite starting by separating determinism / free will and free will / moral responsibility, I follow the frequent convention of merging them into one. Thus, my stance is that because the world is deterministic, there can't be free will, and thus holding people morally responsible for their actions is not okay (a conclusion described as "deplorable" by one leading philosopher whose thinking we're going to dissect big time). This incompatibilism will be most frequently contrasted with the compatibilist view that while the world is deterministic, there is still free will, and thus holding people morally responsible for their actions is just.

Copyrighted Material

This version of compatibilism has produced numerous papers by philosophers and legal scholars concerning the relevance of neuroscience to

free will. After reading lots of them, I've concluded that they usually boil down to three sentences:

a. Wow, there've been all these cool advances in neuroscience, all reinforcing the conclusion that ours is a deterministic world.

b. Some of those neuroscience findings challenge our notions of agency, moral responsibility, and deservedness so deeply that one must conclude that there is no free will.

c. Nah, it still exists.

Naturally, a lot of time will be spent examining the "nah" part. In doing so, I'll consider only a subset of such compatibilists. Here's a thought experiment for identifying them: In 1848 at a construction site in Vermont, an accident with dynamite hurled a metal rod at high speed into the brain of a worker, Phineas Gage, and out the other side. This destroyed much of Gage's frontal cortex, an area central to executive function, long-term planning, and impulse control. In the aftermath, "Gage was no longer Gage," as stated by one friend. Formerly sober, reliable, and the foreman of his work crew, Gage was now "fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom) . . . obstinate, yet capricious and vacillating," as described by his doctor. Phineas Gage is the textbook case that we are the end products of our material brains. Now, 170 years later, we understand how the unique function of your frontal cortex is the result of your genes, prenatal environment, childhood, and so on (wait for chapter 4).

Now the thought experiment: Raise a compatibilist philosopher from birth in a sealed room where they never learn anything about the brain. Then tell them about Phineas Gage and summarize our current knowledge about the frontal cortex. If their immediate response is "Whatever, there's still free will," I'm not interested in their views. The compatibilist I have in mind is one who then wonders, "OMG, what if I'm completely wrong about free will?," ponders hard for hours or decades, and concludes

Copyrighted Material

that there's still free will, here's why, and it's okay for society to hold people morally responsible for their actions. If a compatibilist has not wrestled through being challenged by knowledge of the biology of who we are, it's not worth the time trying to counter their free-will belief.

GROUND RULES AND DEFINITIONS

What is free will? Groan, we have to start with that, so here comes something totally predictable along the lines of "Different things to different types of thinkers, which gets confusing." Totally uninviting. Nevertheless, we have to start there, followed by "What is determinism?" I'll do my best to mitigate the drag of this.

What Do I Mean by Free Will*?*

People define *free will* differently. Many focus on agency, whether a person can control their actions, act with intent. Other definitions concern whether, when a behavior occurs, the person knows that there are alternatives available. Others are less concerned with what you do than with vetoing what you don't want to do. Here's my take.

Suppose that a man pulls the trigger of a gun. Mechanistically, the muscles in his index finger contracted because they were stimulated by a neuron having an action potential (i.e., being in a particularly excited state). That neuron in turn had its action potential because it was stimulated by the neuron just upstream. Which had its own action potential because of the next neuron upstream. And so on.

Here's the challenge to a free willer: Find me the neuron that started this process in this man's brain, the neuron that had an action potential for no reason, where no neuron spoke to it just before. Then show me that this neuron's actions were not influenced by whether the man was tired, hungry, stressed, or in pain at the time. That nothing about this neuron's function was altered by the sights, sounds, smells, and so on, experienced

Copyrighted Material

by the man in the previous minutes, nor by the levels of any hormones marinating his brain in the previous hours to days, nor whether he had experienced a life-changing event in recent months or years. And show me that this neuron's supposedly freely willed functioning wasn't affected by the man's genes, or by the lifelong changes in regulation of those genes caused by experiences during his childhood. Nor by levels of hormones he was exposed to as a fetus, when that brain was being constructed. Nor by the centuries of history and ecology that shaped the invention of the culture in which he was raised. Show me a neuron being a causeless cause in this total sense. The prominent compatibilist philosopher Alfred Mele of Florida State University emphatically feels that requiring something like that of free will is setting the bar "absurdly high."[6] But this bar is neither absurd nor too high. Show me a neuron (or brain) whose generation of a behavior is independent of the sum of its biological past, and for the purposes of this book, you've demonstrated free will. The point of the first half of this book is to establish that this can't be shown.

What Do I Mean by Determinism*?*

It's virtually required to start this topic with the dead White male Pierre Simon Laplace, the eighteenth-/nineteenth-century French polymath (it's also required that you call him a polymath, as he contributed to mathematics, physics, engineering, astronomy, and philosophy). Laplace provided the canonical claim for all of determinism: If you had a superhuman who knew the location of every particle in the universe at this moment, they'd be able to accurately predict every moment in the future. Moreover, if this superhuman (eventually termed "Laplace's demon") could re-create the exact location of every particle at any point in the past, it would lead to a present identical to our current one. The past and future of the universe are already determined.

Science since Laplace's time shows that he wasn't completely right (proving that Laplace was not a Laplacian demon), but the spirit of his demon lives on. Contemporary views of determinism have to incorporate

Copyrighted Material

the fact that certain types of predictability turn out to be impossible (the subject of chapters 5 and 6) and certain aspects of the universe are actually nondeterministic (chapters 9 and 10).

Moreover, contemporary models of determinism must also accommodate the role played by meta-level consciousness. What do I mean by this? Consider a classic psychology demonstration of people having less freedom in their choices than they assumed.[7] Ask someone to name their favorite detergent, and if you have unconsciously cued them earlier with the word *ocean*, they become more likely to answer, "Tide." As an important measure of where meta-level consciousness comes in, suppose the person realizes what the researcher is up to and, wanting to show that they can't be manipulated, decides that they won't say "Tide," even if it is their favorite. Their freedom has been just as constrained, a point in many of the coming chapters. Similarly, wind up as an adult exactly like your parents or the exact opposite of them, and you are equally unfree—in the latter case, the pull toward adopting their behavior, the ability to consciously recognize that tendency to do that, the mindset to recoil from that with horror and thus do the opposite, are all manifestations of the ways that you became you outside your control.

Finally, any contemporary view of determinism must accommodate a profoundly important point, one that dominates the second half of the book—despite the world being deterministic, things can change. Brains change, behaviors change. We change. And that doesn't counter this being a deterministic world without free will. In fact, the science of change *strengthens* the conclusion; this will come in chapter 12.

With those issues in mind, time to see the version of determinism that this book builds on.

Imagine a university graduation ceremony. Almost always moving, despite the platitudes, the boilerplate, the kitsch. The happiness, the pride. The families whose sacrifices now all seem worth it. The graduates who were the first in their family to finish high school. The ones whose immigrant parents sit there glowing, their saris, dashikis, barongs

Copyrighted Material

broadcasting that their pride in the present isn't at the cost of pride in their past.

And then you notice someone. Amid the family clusters postceremony, the new graduates posing for pictures with Grandma in her wheelchair, the bursts of hugs and laughter, you see the person way in the back, the person who is part of the grounds crew, collecting the garbage from the cans on the perimeter of the event.

Randomly pick any of the graduates. Do some magic so that this garbage collector started life with the graduate's genes. Likewise for getting the womb in which nine months were spent and the lifelong epigenetic consequences of that. Get the graduate's childhood as well—one filled with, say, piano lessons and family game nights, instead of, say, threats of going to bed hungry, becoming homeless, or being deported for lack of papers. Let's go all the way so that, in addition to the garbage collector having gotten all that of the graduate's past, the graduate would have gotten the garbage collector's past. Trade every factor over which they had no control, and you will switch who would be in the graduation robe and who would be hauling garbage cans. This is what I mean by determinism.

AND WHY DOES THIS MATTER?

Because we all know that the graduate and the garbage collector would switch places. And because, nevertheless, we rarely reflect on that sort of fact; we congratulate the graduate on all she's accomplished and move out of the way of the garbage guy without glancing at him.

Copyrighted Material

2

The Final Three Minutes of a Movie

Two men stand by a hangar in a small airfield at night. One is in a police officer's uniform, the other dressed as a civilian. They talk tensely while, in the background, a small plane is taxiing to the runway. Suddenly, a vehicle pulls up and a man in a military uniform gets out. He and the police officer talk tensely; the military man begins to make a phone call; the civilian shoots him, killing him. A vehicle full of police pulls up abruptly, the police emerging rapidly. The police officer speaks to them as they retrieve the body. They depart as abruptly, with the body but not the shooter. The police officer and the civilian watch the plane take off and then walk off together.

What's going on? A criminal act obviously occurred—from the care with which the civilian aimed, he clearly intended to shoot the man. A terrible act, compounded further by the man's remorseless air—this was cold-blooded murder, depraved indifference. It is puzzling, though, that the police officer made no attempt to apprehend him. Possibilities come to mind, none good. Perhaps the officer has been blackmailed by the civilian to look the other way. Maybe all the police who appeared on the scene are corrupt, in the pocket of some drug cartel. Or perhaps the police officer is actually an impostor. One can't be certain, but it's clear

Copyrighted Material

that this was a scene of intent-filled corruption and lawless violence, the police officer and the civilian exemplars of humans at their worst. That's for sure.

Intent features heavily in issues about moral responsibility: Did the person intend to act as she did? When exactly was the intent formed? Did she know that she could have done otherwise? Did she feel a sense of ownership of her intent? These are pivotal issues to philosophers, legal scholars, psychologists, and neurobiologists. In fact, a huge percentage of the research done concerning the free-will debate revolves around intent, often microscopically examining the role of intent in the seconds before a behavior happens. Entire conferences, edited volumes, careers, have been spent on those few seconds, and in many ways, this focus is at the heart of arguments supporting compatibilism; this is because all the careful, nuanced, clever experiments done on the subject collectively fail to falsify free will. After reviewing these findings, the purpose of this chapter is to show how, nevertheless, all this is ultimately irrelevant to deciding that there's no free will. This is because this approach misses 99 percent of the story by not asking the key question: *And where did that intent come from in the first place?* This is so important because, as we will see, while it sure may seem at times that we are free do as we intend, we are never free to intend what we intend. Maintaining belief in free will by failing to ask that question can be heartless and immoral and is as myopic as believing that all you need to know to assess a movie is to watch its final three minutes. Without that larger perspective, understanding the features and consequences of intent doesn't amount to a hill of beans.

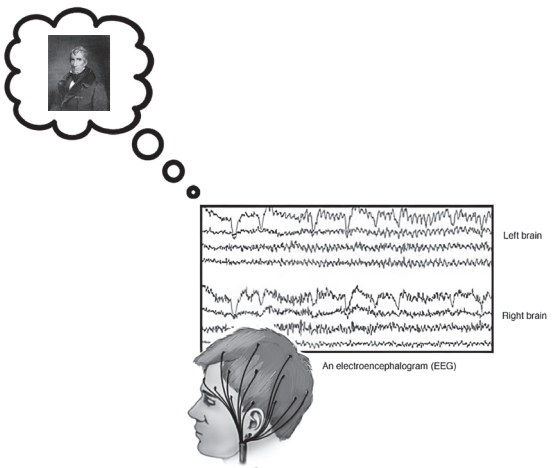
THREE HUNDRED MILLISECONDS

Let's start off with William Henry Harrison, ninth president of the United States, remembered only for idiotically insisting on giving a record-long two-hour inauguration speech in the freezing cold in January 1841,

Copyrighted Material

without coat or hat; he caught pneumonia and died a month later, the first president to die in office and the shortest presidential term.*,[1]

With that in place, think about William Henry Harrison. But first, we're going to stick electrodes all over your scalp for an electroencephalogram (EEG), to observe the waves of neuronal excitation generated by your cortex when you're thinking of Bill.



Now *don't* think of Harrison—think about anything else—as we continue recording your EEG. Good, well done. Now don't think about Harrison, but *plan* to think about him whenever you want a little while later,

*Revisionism suggests that rather than at the inauguration, he caught his pneumonia a few weeks later when, again coatless, he went out to buy a cow. But then even more radical revisionism suggests that he didn't die of pneumonia at all but instead from typhoid fever, contracted from the vile, contaminated water available in the White House. This was concluded by writer Jane McHugh and physician Philip Mackowiak, based on the symptoms detailed by Harrison's doctor and the fact that the White House's water supply was just downstream from where "night soil" was dumped. At the time, Washington, DC, was a malarial swamp, its selection having been advocated by powerful Virginians who wanted the capital close to home; this was decided in behind-closed-doors horse-trading between Alexander Hamilton and Virginians Thomas Jefferson and James Madison. "No one really knows how the game is played, the art of the trade, how the sausage gets made," writes noted historian Lin-Manuel Miranda, referring to the mystery of what transpired in those negotiations.

Copyrighted Material

and push this button the instant you do. Oh, also, keep an eye on the second hand on this clock and note when you chose to think about Harrison. We're also going to wire up your hand with recording electrodes to detect precisely when you start the pushing; meanwhile, the EEG will detect when neurons that command those muscles to push the button start to activate. And this is what we find out: those neurons had already activated *before* you thought you were first freely choosing to start pushing the button.

But the experimental design of this study isn't perfect, because of its nonspecificity—we may have just learned what's happening in your brain when it is generically doing something, as opposed to doing this particular something. Let's switch instead to your choosing between doing A and doing B. William Henry Harrison sits down to some typhoid-riddled burgers and fries, and he asks for ketchup. If you decide he would have pronounced it "ketch-up," immediately push this button with your left hand; if it was "cats-up," push this other button with your right. Don't think about his pronunciation of *ketchup* right now; just look at the clock and tell us the instant you chose which button to push. And you get the same answer—the neurons responsible for whichever hand pushes the button activate before you consciously formed your choice.

Let's do something fancier now than looking at brain waves, since EEG reflects the activity of hundreds of millions of neurons at a time, making it hard to know what's happening in particular brain regions. Thanks to a grant from the WHH Foundation, we've bought a neuroimaging system and will do functional magnetic resonance imaging (fMRI) of your brain while you do the task—this will tell us about activity in each individual brain region at the same time. The results show clearly, once again, that particular regions have "decided" which button to push before you believe you consciously and freely chose. Up to ten seconds before, in fact.

Eh, forget about fMRI and the images it produces, where a single pixel's signal reflects the activity of about half a million neurons. Instead, we're going to drill holes in your head and then stick electrodes into your brain to monitor the activity of individual neurons; using this approach,

Copyrighted Material

once again, we can tell if you'll go for "ketch-up" or "cats-up" from the activity of neurons *before* you believe you decided.

These are the basic approaches and findings in a monumental series of studies that have produced a monumental shitstorm as to whether they demonstrate that free will is a myth. These are the core findings in virtually every debate about what neuroscience can tell us on the subject. And I think that at the end of the day, these studies are irrelevant.

It began with Benjamin Libet, a neuroscientist at the University of California at San Francisco, in a 1983 study so provocative that at least one philosopher refers to it as "infamous," there are conferences held about it, and scientists are described as doing "Libet-style studies."*,[2]

We know the experimental setup. Here's a button. Push it whenever you want. Don't think about it beforehand; look at this fancy clock that makes it easy to detect fractions of a second and tell us when you decided to push the button, that moment of conscious awareness when you freely made your decision.† Meanwhile, we'll be collecting EEG data from you and monitoring exactly when your finger starts moving.

Out of this came the basic findings: people reported that they decided to push the button about two hundred milliseconds—two tenths of a second—before their finger started moving. There was also a distinctive EEG pattern, called a readiness potential, when people prepared to move; this emanated from a part of the brain called the SMA (supplementary motor area), which sends projections down the spine, stimulating muscle movement. But here's the crazy thing: the readiness potential, the evidence that the brain had committed to pushing the button, occurred about three hundred milliseconds *before* people believed they had decided to

*As a point applying to virtually every scientific finding that I'll discuss in the rest of the book, when I say, "work done by John/Jane Doe," it actually means work done by that lead scientist along with a team of collaborators. As an equally important point (that I'll reiterate in various places, because it can't be mentioned too often), when I say, "Scientists showed that when they'd do this or that, people would do X," I mean that *on average*, people responded this way. There are always exceptions, who are often the most interesting.

†In the Libet literature, this point where people thought that they had decided came to be called "W," for the point where they first consciously wished to do something. I'm avoiding using that term, to minimize jargon.

Copyrighted Material

push the button. That sense of freely choosing is just a post hoc illusion, a false sense of agency.

This is the observation that started it all. Read technical papers on biology and free will, and in 99.9 percent of them, *Libet* will appear, usually by the second paragraph. Ditto for articles in the lay press—"Scientist Proves There Is No Free Will; Your Brain Decides Before You Think You Did."* It inspired scads of follow-up research and theorizing; people are still doing studies directly inspired by Libet nearly forty years after his 1983 publication. For example, there's a 2020 paper entitled "Libet's Intention Reports Are Invalid."[3] Having your work be important enough that decades later, people are still trash-talking it is immortality for a scientist.

The basic Libet finding that you're kidding yourself if you think you made a decision when it feels like you did has been replicated. Neuroscientist Patrick Haggard of University College London had subjects choose between two buttons—choosing to do A versus B, rather than choosing to do something versus not. This suggested the same conclusion that the brain has seemingly decided before you think you did.[4]

These findings ushered in Libet 2.0, the work of John-Dylan Haynes and colleagues at Humboldt University in Germany. It was twenty-five years later, with fMRIs available; everything else was the same. Once again, people's sense of conscious choice came about two hundred milliseconds before the muscles started moving. Most important, the study replicated the conclusion from Libet, fleshing it out further.† With fMRI, Haynes was able to spot the which-button decision even farther up in the brain's chain of command, in the prefrontal cortex (PFC). This made sense, as the PFC is where executive decisions are made. (When the PFC, along with the rest of the frontal cortex, is destroyed, à la Gage, one makes

*One paper analyzes the reporting of Libet in the lay press. Eleven percent of the headlines said free will had been disproved; 11 percent said the opposite; many articles were wildly inaccurate in describing how the experiment was done (e.g., saying that it was the researcher who would push the button). And on other fronts, there's even a piece of music called "Libet's Delay." It's moody and so repetitive that I felt a conscious sense of wishing to scream; I can only conclude that it was composed by a deeply depressed AI.

Copyrighted Material

†I'm using "the conclusion from Libet" rather than "Libet's conclusion," in that the latter suggests what Libet himself was thinking about his finding. We'll get to what he thought.

terrible, disinhibited decisions.) To simplify a bit, once having decided, the PFC passes the decision on to the rest of the frontal cortex, which passes it to the premotor cortex, then to the SMA and, a few steps later, on to your muscles.* Supporting the view of Haynes having spotted decision-making farther upstream, the PFC was making its decision up to *ten seconds* before subjects felt they were consciously deciding.†,5

Then Libet 3.0 explored free-will-is-an-illusion down to monitoring the activity of individual neurons. Neuroscientist Itzhak Fried of UCLA worked with patients with intractable epilepsy, unresponsive to antiseizure medications. As a last-ditch effort, neurosurgeons remove the part of the brain where these seizures initiate; with Fried's patients, it was the frontal cortex. One obviously wants to minimize the amount of tissue removed, and in preparation for that, electrodes are implanted in the targeted area prior to the surgery, allowing for monitoring activity there. This provides a fine-grained map of function, telling you what subparts you should avoid removing, if there's any leeway.

So Fried would have the subjects do a Libet-style task while electrodes in their frontal cortex detected when particular neurons there activated. Same punch line: some neurons activated in preparation for a particular movement decision seconds before subjects claimed they had consciously decided. In fascinating related studies, he has shown that neurons in the hippocampus that code for a specific episodic memory activate one to two seconds before the person becomes aware of freely recalling that memory.[6]

Thus, three different techniques, monitoring the activity of hundreds of millions of neurons down to single neurons, all show that at the moment when we believe that we are consciously and freely choosing to do something, the neurobiological die has already been cast. That sense of conscious intent is an irrelevant afterthought.

This conclusion is reinforced by studies showing how malleable the

*One neuroscientist aptly describes the SMA as the "gateway" by which the PFC talks to your muscles.

†Haynes and colleagues have since identified the exact subregion of the PFC involved. They also implicated an additional brain region, the parietal cortex, as part of the decision-making process.

Copyrighted Material

sense of intent and agency is. Back to the basic Libet paradigm; this time, pushing a button caused a bell to ring, and the researchers would vary how long of a fraction-of-a-second time delay there'd be between the pushing and the ringing. When the bell ringing was delayed, subjects reported their intent to push the button coming a bit later than usual—without the readiness potential or actual movement changing. Another study showed that if you feel happy, you perceive that conscious sense of choice sooner than if you're unhappy, showing how our conscious sense of choosing can be fickle and subjective.[7]

Other studies of people undergoing neurosurgery for intractable epilepsy, meanwhile, showed that the sense of intentional movement and actual movement can be separated. Stimulate an additional brain region relevant to decision-making,* and people would claim they had just moved voluntarily—without so much as having tensed a muscle. Stimulate the pre-SMA instead, and people would move their finger while claiming that they hadn't.[8]

One neurological disorder reinforces these findings. Stroke damage to part of the SMA produces "anarchic hand syndrome," where the hand controlled by that side of the SMA† acts against the person's will (e.g., grabbing food from someone else's plate); sufferers even restrain their anarchic hand with their other one.‡ This suggests that the SMA keeps volition on task, binding "intention to action," all before the person believes they've formed that intention.[9]

Psychology studies also show how the sense of agency can be illusory.

*The parietal cortex, mentioned a few footnotes back.

†As a technical detail completely unrelated to any of this, the right half (hemisphere) of the brain regulates movements in the left half of the body; the left hemisphere the reverse.

‡Anarchic hand syndrome, and the closely related "alien hand syndrome," is sometimes called "Dr. Strangelove syndrome"—for the titular character in the 1964 Stanley Kubrick movie. Strangelove was mostly modeled after rocket scientist Wernher von Braun, who went from faithfully serving his Nazi masters during World War II to serving his American ones after; turns out he was a patriotic American all along, that whole Nazi thing just a misunderstanding. Strangelove, wheelchair bound after a stroke, has anarchic hand syndrome, his hand constantly trying to give a Nazi salute to his American overlords. Stanley Kubrick, the famed director of the movie, also incorporated elements of John von Neumann, Herman Kahn, and Edward Teller into Strangelove (but not, despite urban legends, Henry Kissinger).

Copyrighted Material

In one study, pushing a button would be followed immediately by a light going on . . . some of the time. The percentage of time the light would go on was varied; subjects were then asked how much control they felt they had over the light. People consistently overestimate how reliably the light occurs, feeling that they control it.* In another study, subjects believed they were voluntarily choosing which hand to use in pushing a button. Unbeknownst to them, hand choice was being controlled by transcranial magnetic stimulation† of their motor cortex; nonetheless, subjects perceived themselves as controlling their decisions. Meanwhile, other studies used manipulations straight out of the playbook of magicians and mentalists, with subjects claiming agency over events that were actually foregone and out of their control.[10]

If you do X and this is followed by Y, what increases the odds of your feeling like you caused Y? Psychologist Daniel Wegner of Harvard, a key contributor in this area, identified three logical variables. One is *priority*—the shorter the delay between X and Y, the more readily we have an illusory sense of will. There are also *consistency* and *exclusivity*—how consistently Y happens after you've done X, and how often Y happens in the absence of X. The more of the former and the less of the latter, the stronger the illusion.[11]

Collectively, what does this Libetian literature, starting with Libet, show? That we can have an illusory sense of agency, where our sense of freely, consciously choosing to act can be disconnected from reality;‡ we can be manipulated as to when we first feel a sense of conscious control;

*Interestingly, people with depression are resistant to being tricked into this sense of "illusory will." This will be returned to in the final chapter.

†In TMS, an electromagnetic coil is placed on the scalp and used to activate or inactivate the patch of cortex just beneath (I had that done to me once, with the colleague controlling when I bent my index finger; it felt beyond creepy). How's this for a finding whose implications resonate through this book? TMS can be used to alter people's judgments of the moral appropriateness of a behavior.

‡Although, in response to this, philosopher Peter Tse of Dartmouth writes, "Just as the existence of visual illusions does not prove that all vision is illusory, the existence of illusions of conscious agency does not prove that conscious operations cannot be causal of action in certain cases."

Copyrighted Material

most of all, this sense of agency comes after the brain has already committed to an action. Free will is a myth.[12]

Surprise!, people have been screaming at each other about these conclusions ever since, incompatibilists perpetually citing Libet and his descendants, and compatibilists being scornful shade throwers about the entire literature. It didn't take long to start. Two years after his landmark paper, Libet published a review in a peer-commentary journal (where someone presents a theoretical paper on a controversial topic, followed by short commentaries by the scientist's friends and enemies); commentators beating on Libet accused him of "egregious errors," overlooking "fundamental measurement concepts," conceptual unsophistication ("Pardon, your dualism is showing," accused one critic), and having an unscientific faith in the accuracy of his timing measurements (sarcastically proclaiming Libet as practicing "chronotheology").[13]

The criticisms of the work of Libet, Haynes, Fried, Wegner, and friends continue unabated. Some focus on minutiae like the limitations of using EEGs, fMRI, and single-neuron recordings, or the pitfalls inherent in subjects self-reporting most anything. But most criticisms are more conceptual and collectively show that rumors of Libetianism killing free will are exaggerated. These are worth detailing.

YOU GUYS PROCLAIM THE DEATH OF FREE WILL, BASED ON SPONTANEOUS FINGER MOVEMENTS?

The Libetian literature is built around people spontaneously deciding to do something. In the view of Manuel Vargas, free will revolves around being future oriented, enduring an immediate cost for a long-term goal, and thus "Libet's experiment insisted on a purely immediate, impulsive action—which is precisely not what free will is for."[14]

Copyrighted Material

Moreover, what was being spontaneously decided was to push a button,

and this bears little resemblance to whether we have free will concerning our beliefs and values or our most consequential actions. In the words of psychologist Uri Maoz of Chapman University, this is a contrast between "picking" and "choosing"—Libet is about picking which box of Cheerios to take off the supermarket shelf, not about choosing something major. Dartmouth philosopher Adina Roskies, for example, views Libet-world picking as a caricature of real choice, dwarfed even by the complexity of deciding between tea and coffee.*[15]

Does the Libet finding apply to something more interesting than button pushing? Fried replicated the Libet effect when subjects in a driving simulator chose between turning left and turning right. Another study merged neuroscience with getting out of the lab on a sunny day, checking for the Libet phenomenon in subjects just before they bungee-jumped. Did the neuroscientists, clutching their equipment, jump too? No, a wireless EEG device was strapped to the jumpers' heads, making them look like Martians persuaded to bungee-jump by frat bros after some beer pong. Results? Replication of Libet, where a readiness potential preceded the subjects' believing they had decided to jump.[16]

To which the compatibilists replied, This is still totally artificial—choosing when to leap into an abyss or whether to turn left or right in a driving simulator tells us nothing about our free will in choosing between, say, becoming a nudist versus a Buddhist, or becoming an algologist versus an allergologist. This criticism was backed by a particularly elegant study. In the first situation, subjects would be presented with two buttons and told that each represented a particular charity; press one of the buttons and that charity will be sent a thousand dollars. Second version: two buttons, two charities, push whichever button you feel like, each charity is getting five hundred dollars. The brain was commanding the same movement in both scenarios, but the choice in the first one was highly consequential, while that in the second was as arbitrary as the one in the Libet

*While usually classified as a philosopher, Roskies leaves the rest of us pikers in the dust by having a PhD in neuroscience, in addition to her philosophy PhD.

Copyrighted Material

study. The boring, arbitrary situation evoked the usual readiness potential before there was a sense of conscious decision; the consequential one didn't. In other words, Libet doesn't tell us anything about free will worth wanting. In the wonderfully sarcastic words of one leading compatibilist, the take-home message of this entire literature is "Don't play rock paper scissors for money [with one of these free will skeptic researchers] if your head is in an fMRI machine."[17]

But then, the revenge of the free will skeptics. Haynes's group brain-imaged subjects participating in a nonmotoric task, choosing whether to add or subtract one number from another; they found a neural signature of decision coming before conscious awareness, but coming from a *different* brain region than the SMA (called the posterior cingulate / precuneus cortex). So maybe the pick-your-charity scientists were just looking in the wrong part of the brain—simple brain regions decide things before you think you've consciously made a simple decision, more complicated regions before you think you've made a complicated choice.[18]

The jury is still out, because the Libetian literature remains almost entirely about spontaneous decisions regarding some fairly simple things. On to the next broad criticism.

60 PERCENT? REALLY?

What does it mean to become aware of a conscious decision? What do "deciding" and "intending" really mean? Again with semantics that aren't just semantic. The philosophers run wild here in subtle ways that leave many neuroscientists (e.g., me) gasping in defanged awe. How long does it take to focus on focusing on the second hand on a clock? In her writing, Roskies emphasizes the difference between conscious intention and consciousness of intention. Alfred Mele speculates that the readiness potential is the time when, in fact, you have legitimately freely chosen, and it then takes a bit of time for you to be consciously aware of your freely willed

Copyrighted Material

choice. Arguing against this, one study showed that at the time of the onset of the readiness potential, rather than thinking about when they were going to move, many subjects were thinking about things like dinner.[19]

Can you decide to decide? Are intending and having an intent the same thing? Libet instructed subjects to note the time when they first became aware of "the subjective experience of 'wanting' or intending to act"—but are "wanting" and "intending" the same? Is it possible to be spontaneous when you've been told to be spontaneous?

As long as we're at it, what actually is a readiness potential? Remarkably, nearly forty years after Libet, a paper can still be entitled "What Is the Readiness Potential?" Could it be deciding-to-do, actual "intention," while the conscious sense of decision is deciding-to-do-now, an "implementation of intention"? Maybe the readiness potential doesn't mean anything—some models suggest that it is just the point where random activity in the SMA passes a detectable threshold. Mele forcefully suggests that the readiness potential is not a decision but an urge, and physicist Susan Pockett and psychologist Suzanne Purdy, both of the University of Auckland, have shown that the readiness potential is less consistent and shorter when subjects are planning to identify when they made a decision, versus when they felt an urge. For others, the readiness potential is the process leading to deciding, not the decision itself. One clever experiment supports this interpretation. In it, subjects were presented four random letters and then instructed to choose one in their minds; sometimes they were then signaled to press a button corresponding to that letter, sometimes not—thus, the same decision-making process occurred in both scenarios, but only one actually produced movement. Crucially, a similar readiness potential occurred in both cases, suggesting, in the words of compatibilist neuroscientist Michael Gazzaniga, that rather than the SMA deciding to enact a movement, it's "warming up for its participation in the dynamic events."[20]

So are readiness potentials and their precursors decisions or urges? A decision is a decision, but an urge is just an increased likelihood of a decision. Does a preconscious signal like a readiness potential ever occur and

Copyrighted Material

despite that, the movement *doesn't* then happen? Does a movement ever occur *without* a preconscious signal preceding it? Combining these two questions, how accurately do these preconscious signals predict actual behavior? Something close to 100 percent accuracy would be a major blow to free-will belief. In contrast, the closer accuracy is to chance (i.e., 50 percent), the less likely it is that the brain "decides" anything before we feel a sense of choosing.

As it turns out, predictability isn't all that great. The original Libet study was done in such a way that it wasn't possible to generate a number for this. However, in the Haynes studies, fMRI images predicted which behavior occurred with only about 60 percent accuracy, almost at the chance level. For Mele, a "60-percent accuracy rate in predicting which button a participant will press next doesn't seem to be much of a threat to free will." In Roskies's words, "All it suggests is that there are some physical factors that influence decision-making." The Fried studies recording from individual neurons pushed accuracy up into the 80 percent range; while certainly better than chance, this sure doesn't constitute a nail in free will's coffin.[21]

Now for the next criticisms.

WHAT IS CONSCIOUSNESS?

Giving this section this ridiculous heading reflects how unenthused I am about having to write this next stretch. I don't understand what consciousness is, can't define it. I can't understand philosophers' writing about it. Or neuroscientists', for that matter, unless it's "consciousness" in the boring neurological sense, like not experiencing consciousness because you're in a coma.*,[22]

Nevertheless, consciousness is central to Libet debates, sometimes, in a fairly heavy-handed way. For example, take Mele, in a book whose title

Copyrighted Material

*Naturally, it turns out that the neurological distinction between consciousness and unconscious is not boring, simple, or dichotomous, but that's another can of worms.

trumpets that he's not pulling any punches—*Free: Why Science Hasn't Disproved Free Will*. In its first paragraph, he writes, "There are two main scientific arguments today against the existence of free will." One arises from social psychologists showing that behavior can be manipulated by factors that we're not aware of—we've seen examples of these. The other is neuroscientists whose "basic claim is that *all our decisions* are made unconsciously and therefore not freely" (my italics). In other words, that consciousness is just an epiphenomenon, an illusory, reconstructive sense of control irrelevant to our actual behavior. This strikes me as an overly dogmatic way of representing just one of many styles of neuroscientific thought on the subject.

The "ooh, you neuroscientists not only eat your dead but also believe all our decisions are unconscious" nyah-nyah matters, because we shouldn't be held morally responsible for our unconscious behaviors (although neuroscientist Michael Shadlen of Columbia University, whose excellent research has informed free-will debates, makes a spirited argument along with Roskies that we should be held morally responsible for even our unconscious acts).[23]

Compatibilists trying to fend off the Libetians often make a last stand with consciousness: Okay, okay, suppose that Libet, Haynes, Fried, and so on really have shown that the brain decides something before we have a sense of having consciously and freely done so. Let's grant the incompatibilists that. But does turning that preconscious decision into actual behavior require that conscious sense of agency? Because if it does, rather than bypassing consciousness as an irrelevancy, free will can't be ruled out.*

As we saw, knowing what a brain's preconscious decision was moderately predicts whether the behavior will actually occur. But what about the relationship between the preconscious brain's decision and the sense of

*Note that while related, this is subtly different from the issue of whether the sense of conscious decision-making always occurs with the same time lag after the readiness potential; as we saw, the timing of that sense of agency can be manipulated by other factors.

Copyrighted Material

conscious agency—is there ever a readiness potential followed by a behavior without a conscious sense of agency coming in between? One cool study done by Dartmouth neuroscientist Thalia Wheatley and collaborators* shows precisely this—subjects were hypnotized and implanted with a posthypnotic suggestibility that they make a spontaneous Libet-like movement. In this case, when triggered by the cued suggestion, there'd be a readiness potential and the subsequent movement, without conscious awareness in between. Consciousness is an irrelevant hiccup.[24]

Sure, retort compatibilists, this doesn't mean that intentional behavior *always* bypasses consciousness—rejecting free will based on what happens in the posthypnotic brain is kind of flimsy. And there is a higher-order level to this issue, something emphasized by incompatibilist philosopher Gregg Caruso of the State University of New York—you're playing soccer, you have the ball, and you consciously decide that you are going to try to get past this defender, rather than pass the ball off. In the process of then trying to do this, you make a variety of procedural movements that you're not consciously choosing; what does it mean that you have made the explicit choice to let a particular implicit process take over? The debate continues, not just over whether the preconscious requires consciousness as a mediating factor but also over whether both can simultaneously cause a behavior.[25]

Amid these arcana, it's hugely important if the preconscious decision requires consciousness as a mediator. Why? Because during that moment of conscious mediation we should then be expected to be able to *veto* a decision, prevent it from happening. And you can hang moral responsibility on that.[26]

*The study was a collaboration not just between philosophers and neuroscientists but also between people with decidedly incompatibilist stances (Wheatley) and the notable compatibilists Roskies, Tse, and Duke philosopher Walter Sinnott-Armstrong. This is the process of questing for knowledge at its objective best.

Copyrighted Material

FREE WON'T: THE POWER TO VETO

Even if we don't have free will, do we have free won't, the ability to slam our foot on the brake between the moment of that conscious sense of freely choosing to do something and the behavior itself? This is what Libet concluded from his studies. Clearly we have that veto power. Writ small, you're about to reach for more M&M's but stop an instant before. Writ larger, you're about to say something hugely inappropriate and disinhibited but, thank God, you stop yourself as your larynx warms up to doom you.

The basic Libetian findings gave rise to a variety of studies looking at where vetoing actions fits in. Do it or not: once that conscious sense of intent occurs, subjects have the option to stop. Do it now or in a bit: once that conscious sense of intent occurs, immediately push the button or first count to ten. Impose an external veto: In a brain-computer interface study, researchers used a machine learning algorithm that monitored a subject's readiness potential, predicting in real time when the person was about to move; some of the time, the computer would signal the subject to stop the movement in time. Of course, people could generally stop themselves up until a point of no return, which roughly corresponded to when the neurons that send a command directly to muscles were about to fire. As such, a readiness potential doesn't constitute an unstoppable decision, and one would generally look the same whether the subject was definitely going to push a button or there was the possibility of a veto.*,[27]

How does the vetoing work, neurobiologically? Slamming a foot on the brake involved activating neurons just upstream of the SMA.† Libet may have spotted this in a follow-up study examining free won't. Once subjects

*As a fascinating finding in these studies, failing to stop in time activates the anterior cingulate cortex, a brain region associated with subjective feelings of pain; in other words, a few dozen milliseconds is enough time for you to feel like a loser because a computer has gotten a faster draw on you.

†Depending on the study, the pre-SMA, anterior prefrontal cortex, and/or right inferior frontal gyrus. Note that the last two places, logically, implicate the frontal cortex in executive vetoing.

Copyrighted Material

had that conscious sense of intent, they were supposed to veto the action; at that point, the tail end of the readiness potential would lose steam, flatten out.*,[28]

Meanwhile, other studies explored interesting spin-offs of free won't–ness. What's the neurobiology of a gambler on a losing streak who manages to stop gambling, versus one who doesn't?† What happens to free won't when there's alcohol on board? How about kids versus adults? It turns out that kids need to activate more of their frontal cortex than do adults to get the same effectiveness at inhibiting an action.[29]

So what do all these versions of vetoing a behavior in a fraction of a second say about free will? Depends on whom you talk to, naturally. Findings like these have supported a two-stage model about how we are supposedly the captains of our fate, one espoused by the likes of everyone from William James to many contemporary compatibilists. Stage one, the "free" part: your brain spontaneously chooses, amid alternative possibilities, to generate the proclivity toward some action. Stage two, the "will" part, is where you consciously consider this proclivity and either green-light it or free-won't it. As one proponent writes, "Freedom arises from the creative and indeterministic generation of alternative possibilities, which present themselves to the will for evaluation and selection." Or in Mele's words, "even if urges to press are determined by unconscious brain activity, it may be up to the participants whether they act on those urges or not."[30] Thus, "our brains" generate a suggestion, and "we" then judge it; this dualism sets our thinking back centuries.

*The original Libet publication didn't mention anything about flattening out; it was only in a later review that he decided that it occurred. And to be a bit of a killjoy, after looking at the original paper, which had only four subjects, I just don't see it in the shapes of the readiness potentials displayed, and there's no real way to rigorously analyze the shape of each curve, given the data available in the paper; this study happened during a less quantitative, more innocent time for analyzing data.

†Continuing to gamble activated brain regions associated with incentives and reward; in contrast, quitting activated regions related to subjective pain, anxiety, and conflict. This is amazing—continuing to gamble with the possibility of losing is less neurobiologically aversive than quitting and contemplating the possibility that you *would have* won if you hadn't stopped. We're a really screwed-up species.

Copyrighted Material

The alternative conclusion is that free won't is just as suspect as free will, and for the same reasons. Inhibiting a behavior doesn't have fancier neurobiological properties than activating a behavior, and brain circuitry even uses their components interchangeably. For example, sometimes brains do something by activating neuron X, sometimes by inhibiting the neuron that is inhibiting neuron X. Calling the former "free will" and calling the latter "free won't" are equally untenable. This recalls chapter 1's challenge to find a neuron that initiated some act without being influenced by any other neuron or by any prior biological event. Now the challenge is to find a neuron that was equally autonomous in preventing an act. Neither free-will nor free-won't neurons exist.

Having now reviewed these debates, what can we conclude? For Libetians, these studies show that our brains decide to carry out a behavior before we think that we've freely and consciously done so. But given the criticisms that have been raised, I think all that can be concluded is that in some fairly artificial circumstances, certain measures of brain function are moderately predictive of a subsequent behavior. Free will, I believe, survives Libetianism. And yet I think that is irrelevant.

JUST IN CASE YOU THOUGHT THIS WAS ALL ACADEMIC

The debates over Libet and his descendants can be boiled down to a question of intent: When we consciously decide that we intend to do something, has the nervous system already started to act upon that intent, and what does it mean if it has?

A related question is screamingly important in one of the areas where this free-will hubbub is profoundly consequential—in the courtroom. When someone acts in a criminal manner, did they intend to?

Copyrighted Material

By this I'm not suggesting bewigged judges arguing about some low-life's readiness potentials. Instead, the questions that define "intent" are whether a defendant could foresee, without substantial doubt, what was going to happen as a result of their action or inaction, and whether they were okay with that outcome. From that perspective, unless there was intent in that sense, a person shouldn't be convicted of a crime.

Naturally, this generates complex questions. For example, should intending to shoot someone but missing count as a lesser crime than shooting successfully? Should driving with a blood alcohol level in the range that impairs control of a car count as less of a transgression if you lucked out and happened not to kill a pedestrian than if you did (an issue that Oxford philosopher Neil Levy has explored with the concept of "moral luck")?[31]

As another wrinkle, the legal field distinguishes between *general* and *specific* intent. The former is about intending to commit a crime, whereas the latter is intending to commit a crime as well as intending a specific consequence; the charge of the latter is definitely more serious than the former.

Another issue that can come up is deciding whether someone acted intentionally out of fear or anger, with fear (especially when reasonable) seen as more mitigating; trust me, if the jury consisted of neuroscientists, they'd deliberate for eternity trying to decide which emotion was going on. How about if someone intended to do something criminal but instead unintentionally did something else criminal?

An issue that we all recognize is how long before a behavior the intent was formed. This is the world of premeditation, the difference between, say, a crime of passion with a few milliseconds of intent versus an action long planned. It is pretty unclear legally exactly how long one needs to meditate upon an intended act for it to count as premeditated. As an example of this lack of clarity, I once was a teaching witness in a trial where a pivotal issue was whether eight seconds (as recorded by a CCTV camera) is enough time for someone in a life-threatening circumstance

Copyrighted Material

to *pre*meditate a murder. (My two cents was that under the circumstances involved, eight seconds not only wasn't enough time for a brain to do premeditated thinking, it wasn't enough time for it to do any thinking, and free won't–ness was an irrelevant concept; the jury heartily disagreed.)

Then there are questions that can be at the core of war crime trials. What kind of threat is needed for someone's criminality to count as coerced? What about agreeing to do something with criminal intent while knowing that if you refused, someone else would do it immediately and more brutally? Taking things even further, what should be done with someone who intentionally chose to commit a crime, not knowing that they would have been forced to commit that act if they had tried to do otherwise?*,[32]

At this juncture, we appear to have two wildly different realms of thinking about agency and responsibility—people arguing about the supplementary motor area in neurophilosophy conferences and prosecutors and public defenders jousting in courtrooms. Yet they share something that potentially strikes a blow against free-will skepticism:

> Suppose it turns out that our sense of conscious decision-making doesn't actually come after things like readiness potentials, that activity in the SMA, the prefrontal cortex, the parietal cortex, wherever, is

*It seems intuitive that someone should be punished if they thought they had willingly chosen to do something illegal without knowing that they actually didn't have a choice. The late Princeton philosopher Harry Frankfurt has taken the implications of this intuition in a particular compatibilist direction. Step 1: Incompatibilists say that if the world is deterministic, there shouldn't be moral responsibility. Step 2: Consider someone choosing to do something, not knowing that they would have been coerced if they hadn't. Step 3: Therefore, this would be a deterministic world, in that the person didn't actually have the option of doing otherwise . . . yet our intuitions are to hold him morally responsible, perceiving him as having had free will. Huzzah, we've thus just proven that free will and moral responsibility are compatible with determinism. I feel bad saying this because Frankfurt looks cherubic in his pictures, but this seems like more than a bit of sophistry and sure doesn't represent the Demise of Incompatibilism. Moreover, I get the sense from friends in the know that while Frankfurt is enormously influential in some corners of legal philosophy, millennia go by without these "Frankfurt counterexamples" being relevant in an actual courtroom; it is unlikely for there to be scenarios where "the defendant chose to slap the Oscar host across the face, not aware that if he had not chosen to do so, he would nevertheless have been forced to."

Copyrighted Material

never better than only moderately predicting behavior, and only for the likes of pushing buttons. You sure can't say free will is dead based on that.

Likewise, suppose a defendant says, "I did it. I knew there were other things I could do, but I intended to do it, planned it in advance. I not only knew that X could have been the outcome, I wanted that to happen." Good luck convincing someone that the defendant lacked free will.

But the point of this chapter is that even if either or both of these are the case, I still think that free will doesn't exist. To appreciate why, time for a Libet-style thought experiment.

THE DEATH OF FREE WILL IN THE SHADOW OF INTENT

You have a friend doing research for her doctorate in neurophilosophy, and she asks you to be a test subject. Sure. She's upbeat because she's figured out how to both get another data point for her study *and* simultaneously accomplish something else that she's keen on—win-win. It involves ambulatory EEG, out of the lab, like in the bungee jumping study. You're out there now, wired up with the leads, electromyography being done on your hand, a clock in view.

As with the classic Libet, the motoric action involved is to move your index finger. Hey, aren't we decades past that sort of really artificial scenario? Fortunately, the study is more sophisticated than that, thanks to your friend's careful experimental design—you'll be making a simple movement, but with a nonsimple consequence. Don't plan ahead to make this movement, you're told, do it spontaneously, and note on the clock what time it is when you first consciously intend to. All set? Now, when you feel like it, pull a trigger and kill this person.

Copyrighted Material

Maybe the person is an enemy of the Fatherland, a terrorist blowing up

bridges in one of the gloriously occupied colonies. Maybe it's the person behind the cash register in the liquor store you're robbing. Maybe they're a terminally ill loved one in unspeakable pain, begging you to do this. Maybe it's someone who is about to harm a child; maybe it is the infant Hitler, cooing in his crib.

You are free to choose not to shoot. You're disillusioned with the regime's brutality and refuse; you think killing the clerk ups the ante too much if you're caught; despite your loved one begging, you just can't do it. Or maybe you're Humphrey Bogart, your friend is Claude Rains, you're confusing reality with story line and figure that if you let Major Strasser escape, the story doesn't end and you'll get to star in a sequel to *Casablanca*.*

But suppose you have to pull the trigger or else there'll be no readiness potential to detect and your friend's research will be slowed down. Nonetheless, you still have options. You can shoot the person. You can shoot but intentionally miss. You can shoot yourself rather than comply.† As a major plot twist, you can shoot your friend.

It makes intuitive sense that if you want to understand what you wind up doing with your index finger on that trigger, that you should explore Libetian concerns, studying particular neurons and particular milliseconds in order to understand the instant you feel you have chosen to do something, the instant your brain has committed to that action, and whether those two things are the same. But here's why these Libetian debates, as well as a criminal justice system that cares only about whether someone's actions are intentional, are irrelevant to thinking about free will. As first aired at the beginning of this chapter, that is because neither asks a question central to every page of this book: *Where did that intent come from in the first place?*

If you don't ask that question, you've restricted yourself to a domain of

*Aha!

†The Dalai Lama was once asked what he would do in the "runaway trolley" problem (a trolley whose brakes have failed is hurtling down the tracks, about to kill five people; is it okay to push someone in front of the trolley, intentionally killing them but preventing the deaths of five?); he said he would throw himself in front of the trolley.

Copyrighted Material