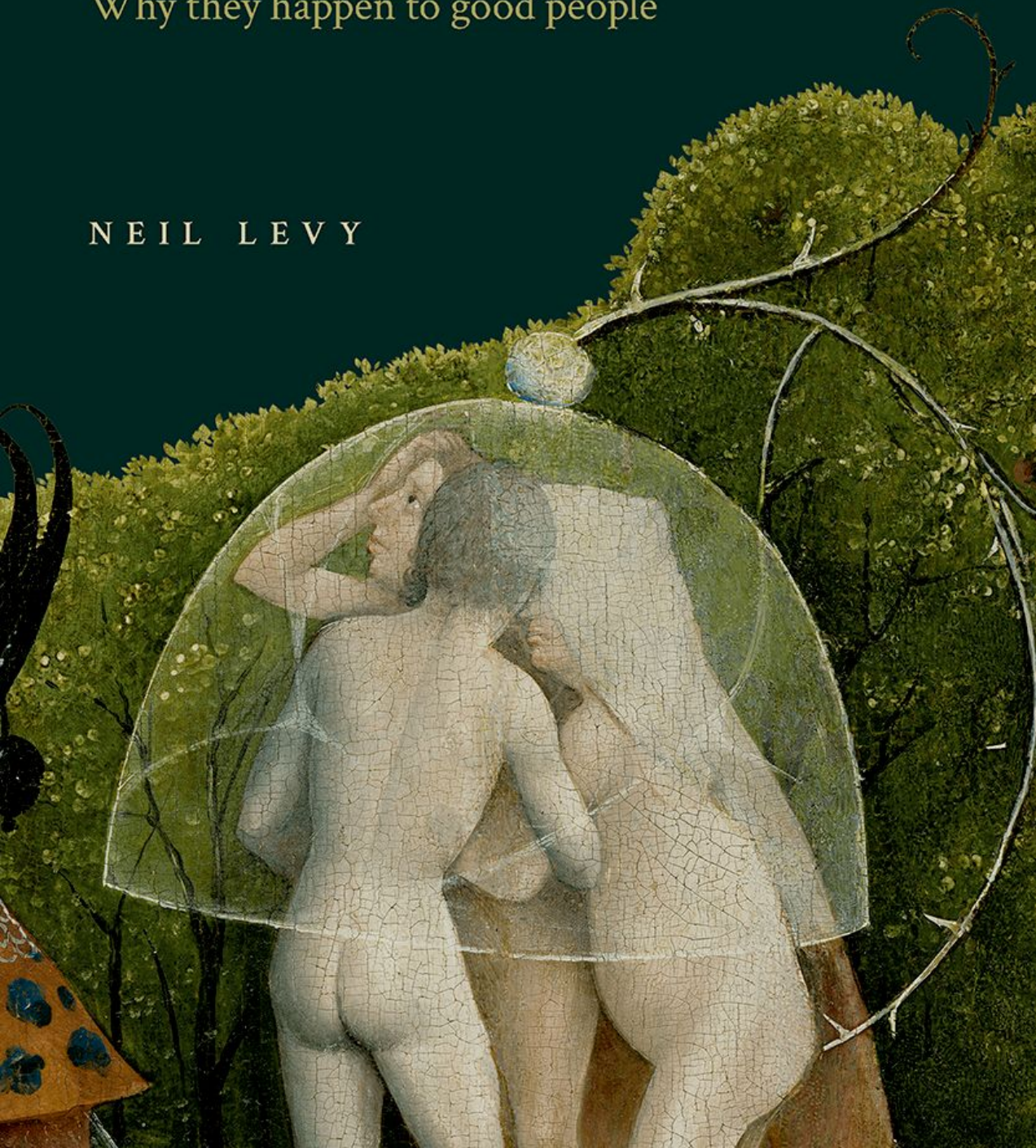
OXFORD

Bad Beliefs

Why they happen to good people

NEIL LEVY



Bad Beliefs

Bad Beliefs

Why They Happen to Good People

NEIL LEVY

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2022

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2021940779

ISBN 978–0–19–289532–5

DOI: 10.1093/oso/9780192895325.001.0001

Printed and bound in the UK by
TJ Books Limited

Contents

Acknowledgments

Eighteen months ago, I started writing a very different book. That book, which drew heavily on previously published work of mine, explained bad belief formation as the product of the ways in which we fall short of rationality. During the months of lockdown, first in the UK and then in Australia, I came to a very different view. According to the account I now defend, bad beliefs are produced by rational agents responding appropriately to evidence.

Since my view is now so different from the one I earlier espoused, I've departed much further from previously published work than I had expected. Still, in retrospect that work can be seen as groping toward my current position, and I have drawn from previously published papers where they were helpful. While most have been heavily revised, I have drawn from the following papers:

Levy, N., & Alfano, M. (2019). "Knowledge from vice: Deeply social epistemology," *Mind*, 515: 887–915.

Levy, N. (2019). "Nudge, nudge, wink, wink: Nudging is giving reasons," *Ergo: An Open Access Journal of Philosophy*, 6(10). http://dx.doi.org/10.3998/ergo.12405314.0006.010.

Levy, N. (2019). "Due deference to denialism: Explaining ordinary people's rejection of established scientific findings | SpringerLink," *Synthese*, 196/1: 313–27.

Levy, N. (2018). "Taking responsibility for health in an epistemically polluted environment," *Theoretical Medicine and Bioethics*, 39: 123–41.

Levy, N. (2016). "The earthling's secret weapon: Cumulative culture and the singularity," *Science, Religion and Culture*, 3: 19–30.

Levy, N. (2021). "Not so hypocritical after all." De Smedt J. & De Cruz H. (eds) *Empirically Engaged Evolutionary Ethics*, pp. 41–61. Springer.

Levy, N., & Ross, R. (2021). "The cognitive science of fake news." Hannon M. & De Ridder J. (eds) *The Routledge Handbook of Political Epistemology*, pp. 181–191. Routledge.

Levy, N. (forthcoming). "Arrogance and servility online." Alfano, M., Lynch, M.P. & Tanesini, A. (eds) *Routledge Handbook of the Philosophy of Humility*. Routledge.

A number of people provided generous comments on part or all of the manuscript. I'm particularly grateful to Mark Alfano, Michael Brownstein, Allen Buchanan, Brian Earp, Gabriel De Marco, Konrad Talmont-Kominski, Neil Van Leeuwen, and Robert Ross for their comments. I'd also like to thank two anonymous reviewers for OUP for helpful feedback. I am especially grateful to the Wellcome Trust (WT104848/Z/14/Z) and the Australian Research Council (DP180102384) for their support.

Preface: Rational Social Animals Go Wild

We live in polarized times. We are politically and ideologically divided, and our divisions seem to reflect, or be constituted by, divergent *beliefs.* As I write, the US is split into two camps that seem to have wildly divergent beliefs over the efficacy of facemasks. For one side, they are sensible precautions against the spread of COVID-19; for the other, they represent an outrageous infringement of civil liberties. The value of facemasks is just one topic that divides right and left over the virus and its risks. The two sides diverge on the origin of the virus, on the advisability of lockdowns and on the harmfulness of the disease, among other topics.

COVID-19 is just the latest front in an ongoing conflict that pits beliefs against beliefs. While the science of COVID-19 is young, and there's room to doubt whether *either* side has a right to much confidence in its views, on many of the issues that divide us the evidence is overwhelmingly on one side. Bad beliefs—beliefs that appear to be wildly at variance with the great preponderance of evidence—seem to animate opposition to vaccination and to the teaching of evolution. Much more consequentially—quite probably catastrophically—bad beliefs have played a central role in the world's failure to tackle climate change. Until very recently, US policy on climate change was decided by a president who thinks that global warming is a hoax. Together with Australia, Brazil, and other countries governed by climate change deniers, the US was able to stymie international efforts at tackling the problem. The world is already paying a heavy price.

This is a book about beliefs, good and bad, about how they are generated and how they might best be improved. Epistemology, the subdiscipline of philosophy concerned with beliefs and their justification, is ancient. Up to quite recently, however, modern epistemology was focused on theoretical questions; in particular, on the analysis of knowledge. It wasn't much concerned with the practical questions that will be

my focus. My concern isn't with the analysis of knowledge. Rather, it is with how knowledge is acquired and what factors lead to good and bad beliefs. Correlatively, my exemplars of belief will not be the uncontroversial cases that feature in a great deal of contemporary discussion; cases in which, say, one agent believes that another owns a car of a particular make, or that there's a barn hereabouts. Instead, my exemplars will be cases that are controversial but *shouldn't* be: beliefs about anthropogenic climate change, evolution, and the safety and efficacy of vaccines. These examples are chosen because there's an expert consensus on these issues, but many people reject the expert view. Are they rational in doing so? What explains their dissent? Should we attempt to change their minds, and if so, how should we do so? I'll defend some controversial answers to these questions about controversial beliefs.

Philosophers love definitions, so before I go on let me say a few words about what I mean by "bad beliefs." There are lots of ways in which beliefs might be bad. A belief might be morally bad (racist beliefs are bad in that kind of way). I'm not concerned with moral badness, but with epistemic badness; that is badness in the belief's relationship to evidence and to the world it aims to reflect. Epistemic badness itself comes in a variety of forms. One way in which a belief can be epistemically bad is by being false. My primary examples of bad belief are false: climate change denial, anti-vaxxer beliefs, creationism, and so on. But not all false beliefs are bad beliefs, in the sense I'm concerned with. I'm an atheist: I don't believe that any religion is true. But I don't think theists are bad believers in my sense. They're not bad believers because I think that religious belief can be rational: a thoughtful person who is familiar with the evidence for and against the existence of God can reasonably conclude that God exists.

A bad belief, in my sense, is therefore not (necessarily) a false belief, but an unjustified belief (a bad belief could even be a true belief, if the totality of evidence is misleading). That still isn't enough to pin down the kind of belief I'm concerned with, though: there are different ways in which beliefs can be unjustified, and I'm concerned with only one. A belief might be unjustified because it's not supported by the evidence available to the believer herself or because it's not supported by the totality of the evidence. Obviously, these things can come apart: a detective

might conclude her suspect is guilty of the crime based on nothing but prejudice and therefore believe badly (in one way) even if all the available evidence (including, say, the forensic report that she hasn't yet read) actually supports the suspect's guilt. Bad beliefs, in my sense, are not those that are subjectively unjustified in this kind of way, though.

A bad belief, in my sense, is a belief that (a) conflicts with the beliefs held by the *relevant epistemic authorities* and (b) held despite the widespread public availability either of the evidence that supports more accurate beliefs or of the knowledge that the relevant authorities believe as they do. The "relevant epistemic authorities" are those people and institutions that are widely recognized as being in the best position to answer questions in the domain: scientists are the relevant epistemic authorities when it comes to evolution; historians the relevant epistemic authorities on the Holocaust; and so on. I don't intend this characterization as a *definition* of bad beliefs (philosophers, save your counterexamples). It's meant to pick out a set of beliefs without begging some questions I want to leave open, such as the question whether bad believers can be subjectively justified in holding their beliefs. My aim is to explain why people come to hold beliefs that are bad in *that* kind of way. Why do they reject climate change, in defiance of the scientific authorities? Why do they reject vaccines, in defiance of the medical profession? And so on.

There are already many books and papers which aim to answer this question or questions that overlap very considerably with this one. Many of these books and papers argue that bad beliefs are explained (in important part) by the ways in which we humans are supposed to depart from some ideal of rational deliberation. On these kinds of views, bad beliefs are explained by a range of irrational (or *a*rational) psychological dispositions characteristic of human beings (albeit perhaps more pronounced on one side of politics than the other). According to views like this, people reject science (say) due to a need for stability, or out of an irrational respect for authority figures or (in more scientific language) due to the confirmation bias or a need for cognitive closure. These biases or dispositions are irrational or arational, though *having* them might itself be rational. The world is complex and time is short; often we must make decisions on the basis of limited information and before we can

properly think things through. It is often better to rely on processes that tend to get things right most of the time, even if they are not themselves rational. On a now standard story, we have evolved to think rationally only under certain conditions: when time permits and resources are plentiful and we're motivated to draw on those resources. These influential views explain bad beliefs by our tendency to make decisions using heuristics and shortcuts, reliance on which is inevitable and adaptive.

This book defends a very different view. It argues that bad beliefs are (in central cases) the product of genuinely and wholly *rational* processes. These processes are rational in the sense that they respond appropriately to evidence, *as* evidence. To say that they're rational is not to say that they always get things right (though given the link between evidence and getting things right, rational mechanisms tend to get things right). If the evidence is misleading, rational processes go wrong. If I get lost in a foreign city because the map I got from the tourist office has been tampered with by pranksters, my confusion is not due to any rational failing on my part. I accessed and processed information in a way that's beyond criticism (assuming there are no grounds for me to suspect tampering). The problem arises in the *inputs* into my navigational reasoning, not my reasoning itself. Analogously, I'll suggest that bad beliefs tend to arise—very significantly, at any rate—through the rational reasoning processes of those who end up believing badly. Just as I might find myself lost because someone tampered with the inputs into my navigation, so people end up believing badly because their epistemic environment has been manipulated.

Another way in which this book departs from more familiar views is that it heavily emphasizes social processes in the generation of knowledge. Knowledge, I'll argue, arises from distributed epistemic labor: epistemic labor distributed across space, time and across agents. Moreover, the knowledge thereby generated often is itself not an individual possession: it is parceled out across multiple agents, and even across the environment. It is *normal* and *rationally appropriate* for agents not to fully understand their own epistemic tools or the role they themselves play in the generation of knowledge, nor even the knowledge thereby generated. This is not just a limitation to which people like you and I are subject, because we're not scientists. Rather, it is the expected

upshot of the way cognitive labor is distributed, and scientists are and must be limited in just the same kind of way.

I'll argue that those who come to hold bad beliefs do so for roughly the same sorts of reasons as those who come to hold good beliefs. It isn't because *they're* irrational and *we're* not. It is largely because *we* defer to reliable sources of evidence and *they* defer to unreliable. This deference, which may be explicit or implicit, is itself rational on both sides. Given that we're epistemically social animals, it's largely through deference that we come to know about the world and generate further knowledge. The processes are much the same in our case and in theirs, and for the most part beyond reproach. Accounting for why some of us go astray in belief formation requires us to understand mechanisms of deference, the features of agents and the world that lead us to trust one source rather than another, and how testimony can be implicit as well as explicit. It also opens us onto the world: it requires us to scrutinize the features of the epistemic landscape and how that landscape can come to be epistemically polluted.

The rationality of bad belief formation has escaped recognition by philosophers, social scientists, and the general public, I suggest, because bad beliefs are so at odds with so much of the evidence. Climate change skeptics have beliefs that are at odds with the record of climate change and with well-established theories about the relationship between CO_2 and temperature. Anti-vaxxers have beliefs about the safety of vaccines that are at odds with the medical literature. And so on. I'll argue that nevertheless these beliefs are not at odds with the *higher-order* evidence. Higher-order evidence is evidence that concerns not the issues about which we're trying to make up our minds, but the reliability of the first-order evidence and how other people are responding to that evidence. Higher-order evidence is genuine evidence, and we rely on it all the time. But philosophers and psychologists overlook its pervasiveness and its significance. Once we come to see the ubiquity of higher-order evidence and the extent to which cognition is reliant on it, we'll be forced to rethink the extent of irrationality in human reasoning.

In effect, the argument I offer from high-order evidence parallels Cecilia Heyes' (2018) argument against nativist accounts of cognition. Nativists appeal to the "poverty of the stimulus" to motivate their

accounts: given that infants receive so little instruction and have so few examples to imitate, the acquisition of species-typical behaviors and capacities must be due to genes and not environment. Heyes argues that this is false: the stimulus is rich, not impoverished. She argues that the infant has ample opportunities for learning. I suggest that an analogous appeal to poverty underwrites arguments for pervasive human irrationality: given how impoverished the evidence for bad beliefs is, something other than rational response to evidence must explain their formation. In the experiments designed to demonstrate irrationality, first-order evidence for the beliefs adopted may be thin, but there's plenty of evidence nevertheless, and our responses are typically sensitive to it, I'll argue.

An account of knowledge production that emphasizes how the epistemic environment is saturated with higher-order evidence yields a distinctive account of how we best improve knowledge production. Bad belief has bad political effects, but it also has causes that are themselves political, in a broad sense of "political." We are (I'll suggest) epistemic individualists: we prize individual cognition and take it to be responsible for the great bulk of our cognitive achievements.[1] This individualism causes us to overlook or underestimate the need to attend to the epistemic environment: to the social mechanisms underlying knowledge production and the social cues that modulate deference. I'll suggest that understanding knowledge and belief requires combating epistemic individualism, and being more attentive to our environment and to the pollutants that have been allowed to accumulate in it.

A focus on the epistemic environment leads to different kinds of remedies for bad beliefs than those suggested by more familiar views. Deficit accounts of bad belief formation (a deficit of knowledge, of motivation, of rationality) suggest remedies that turn on correcting the deficit(s). If the deficit is in information, then we might improve beliefs by broadcasting the truth more widely. If it is in rationality, we might address it

[1] Who is the "we" here? The evidence is strongest with regard to WEIRD people; that is, those who live in the Western, Educated, Industrialized, Rich and Developed world. It is WEIRD people who are the participants in most psychological studies (in most research in most fields), and therefore less is known about other groups. There is some, albeit contested, evidence that other cultures are less individualistic than WEIRD cultures, though as we will see there is evidence that East and South Asians are also epistemic individualists.

through the education system (perhaps by teaching critical thinking). Alternatively, perhaps it would best be addressed by presenting information in a way that minimizes its potential threat to identity or to people's values. At least some of the approaches inspired by these explanations are valuable. What's not to like about improving the education system? I'm confident that some of these initiatives would actually improve people's beliefs to some degree. But I'll suggest that they should not be our sole, or even our main, focus. Rather, we should focus on improving the epistemic environment. That doesn't just mean we need to address what messages are circulating (this isn't the information deficit hypothesis in a new guise). The epistemic environment consists in much more than explicit messages. It consists in agents and institutions as well as messages, and the former may often be more significant than the latter.

Consider, for instance, the cues that we use to decide how much weight to give to testimony. Some of these cues are obvious: for instance, we weigh testimony by those we perceive as expert more heavily than testimony from those we perceive as less expert, and we weigh testimony from multiple sources more heavily than testimony from a lone individual. Only a little less obviously, we weigh testimony by those we perceive as sharing our values more heavily than from those we perceive as malevolent or as ideological opponents. Given these facts, one important way of improving people's beliefs is by way of attending to these kinds of cues. We can improve belief formation through what I will call *epistemic engineering*: the management of the epistemic environment. For instance, we might take care to ensure that people who lack expertise can't easily give themselves an unearned *appearance* of expertise.

Epistemic engineering raises significant ethical issues, of course. Aren't we manipulating others when we engineer the environment in this kind of way? To see the force of the objection, contrast such engineering with more traditional ways of changing minds: by giving reasons and presenting evidence. These more traditional ways are (surely) maximally respectful of agents and their rationality. In contrast, changing the ways in which cues for belief are distributed seems disrespectful at best, perhaps even subversive of agents' autonomy. The worry becomes even more pressing if a great deal of testimony is implicit, delivered not

via assertion but implicated by what is not said, and even by subtle features of the context. Manipulating such features, in the way I advocate, seems to be engaging in *nudging*, and nudging is hugely controversial.

I'll argue that in its typical guises, nudging (and therefore epistemic engineering) is unproblematic. It's controversial only because it's misunderstood. Both philosophers and cognitive scientists typically understand nudging as taking advantage of non-rational mechanisms; it's because nudges bypass genuine reasoning that nudging is controversial. I argue instead that nudges should be understood as implicit testimony. Being guided by a nudge is being guided by testimony, and there's nothing irrational about such guidance (here my defense of the rationality of the processes underlying belief formation becomes relevant to the assessment of the policies aimed at improving it). Hence nudging is usually respectful of agency, and questions concerning manipulation or epistemic paternalism can be set aside.

The Book: A Preview

Darwin called *The Origin of Species* "one long argument." This book, too, is an attempt at one long argument: it's designed to be read through (here, I'm afraid, comparisons with Darwin come to an end). Nevertheless, readers might appreciate a sense of the contents to come.

In Chapter 1, I introduce the topic of belief and belief formation, and set out the case for thinking that the quality of our beliefs is crucial to the quality of our social and political lives. I make this case against the belief skeptics: those who think that the beliefs that agents express play a smaller role in explaining their behavior than we might have thought. I then turn to existing explanations of belief acquisition and update (i.e, how they change over time), drawn from the social sciences. I argue that the influential deficit and motivated cognition accounts fall short of explaining how people come to hold entrenched views that conflict with settled science. In Chapter 2, I turn to a very different body of work in the cognitive sciences: work on cultural evolution. Drawing heavily on the so-called Californian school of researchers, I argue that we owe much of our success at colonizing a dizzying variety of environments to

cumulative culture, which embodies valuable knowledge. This knowledge, I suggest, is deeply social: it's the product of cognition distributed across many agents and across time, and it is never fully grasped by any individual. I then argue that contemporary science does not free us from heavy reliance on socially embedded knowledge production. Rather, if anything, it increases it: science, too, is the product of distributed cognition and individual scientists are never in a position fully to understand their own work.

Chapter 3 turns to distributed cognition in everyday contexts. I argue that the outsourcing of knowledge to others is routine for us. We take ourselves to be epistemically autonomous beings, but we form and update our beliefs very heavily through social referencing (looking to others, especially to those with whom we identify) and deference. These kinds of processes are rational, I'll argue: they're ways of responding appropriately to genuine evidence. They're also highly adaptive, though it's not that fact that makes them rational (it's the other way round). Having established that distributed cognition is more powerful than we tend to think, I turn in Chapters 4 and 5 to the converse question: how successful is individual cognition? In Chapter 4 I argue, contrary to what seems to be the consensus in epistemology and contrary to widespread intuition, that unaided individual cognition is highly unreliable. Without deference (to the right people to the right extent), we're epistemically at sea. In Chapter 5 I argue that we live in a polluted epistemic environment, which ensures that individual cognition fares even worse than it might've done. The focus on individual reasoning has led us to neglect this environment, I'll argue, thereby handing its management over to frauds and merchants of doubt.

In advocating attention to the epistemic environment, I'm advocating nudging, and nudging is highly controversial. In Chapter 6, I address this issue. I argue that nudging is not autonomy-subversive, as is often thought. It's not autonomy-subversive because it relies on mechanisms of deference—the same sorts of mechanisms that in earlier chapters I suggested were rational mechanisms. Nudging is in effect arguing, and being guided by nudges is being guided by (higher-order) evidence. Chapter 6 has an additional aim. Not only does it aim to show that nudges provide higher-order evidence: it also aims to show that the

cultural and social cues that in earlier chapters I argued are essential to human flourishing and to knowledge production themselves work through the provision of higher-order evidence. We orient ourselves and make decisions *centrally* by reference to higher-order evidence.

In a brief concluding chapter, I pull these threads together. Higher-order evidence is genuine evidence: in being guided by it, we're acting and thinking rationally. It follows, I argue, that much of the evidence commonly cited in support of the view that we are pervasively ir- or arational does not in fact support it. We've tended to conclude, on the basis of evidence that we're often responsive to cues and manipulations that don't involve the presentation of first-order evidence, that we're responding arationally (albeit adaptively). But we're more rational than we think: we're social and cultural animals, and we respond to the genuine evidence that our fellows and our cultural environment provide to us. Perhaps we're rational animals after all.

That's the agenda this book will pursue. Let me finish this introduction with a few words on methodology, and the sources of evidence that will guide my argument. I am a philosopher, working in that tradition of post-analytic philosophy that takes the sciences as exemplary (though not, to my mind, exhaustive) of knowledge production (this is a branch of what Eric Schliesser (2019) calls *synthetic philosophy*). The kind of philosophy I aim to engage in develops theories that systematize and interpret evidence from a broad range of sources, but especially from the cognitive sciences: cognitive and social psychology, the cognitive science of religion and work in cultural evolution. I engage in this kind of philosophy—call it naturalistic synthetic philosophy—because I believe it's more likely to generate knowledge about the kinds of questions I am interested in (here) than alternatives. That doesn't mean that I think other ways of doing philosophy are worthless. Far from it: Other approaches may be better for pursuing other valuable ends. Further, other ways of doing philosophy are often relevant to my project, and I'll draw on them when they are.

In particular, I'll draw on work in epistemology. Above, I mentioned that the focus of modern epistemology has been on the analysis of knowledge. But recently there's been a flowering of work in analytic epistemology focused on more practical questions. Analytic philosophy

differs from naturalistic synthetic philosophy in that while the latter takes the sciences as its most important source of evidence, the former relies heavily on the tools of conceptual analysis, the construction of thought experiments and the generation of counterexamples. Analytic epistemology has recently spawned social epistemology: epistemology concerned with the epistemic workings and effects of social interaction and institutions. I've already signaled my indebtedness to social epistemology by referring to testimony above. I'll draw on work in social epistemology and in analytic epistemology more broadly: for instance, work on higher-order evidence and on the epistemic significance of disagreement.

In a recent book, Nathan Ballantyne (2019) describes his project as an exercise in "inclusive regulative epistemology." Regulative epistemology is practical: it aims at guiding belief formation. Ballantyne's work is inclusive because unlike some other regulative epistemologists (he cites Bishop and Trout (2004) and Roberts and Wood (2007)) it does not aim to replace other methods, but instead draws on them. Ballantyne's project is an exercise in inclusive analytic regulative epistemology. Mine– in an even bigger mouthful– is an exercise in inclusive naturalistic synthetic regulative epistemology (don't worry—there'll be no call for me to use this phrase again).

Having situated the project on the philosophical field, let me now say something about its relationship to the cognitive sciences. In recent years, psychology has been rocked by a replication crisis: when experiments have been repeated, researchers have often been unable to reproduce the original findings. For instance, one group attempted to replicate 100 experiments previously published in high-profile journals, but succeeded in replicating only 41 (Open Science Collaboration 2015).[2] This crisis has made some philosophers reluctant to utilize evidence from psychology, and has led others to dismiss the entire field and the philosophy that draws on it. Caution is warranted, but dismissal is not.

[2] It's important to note, however, that these data are difficult to interpret. A failure to reach significance—the criterion they used for successful replication—might in many cases be due to the power of the replication attempt. It's also important to note that failures of replication may occur for many reasons, even when the underlying effect is real. See Earp (2016) for discussion of how we should understand this project.

While it's not strictly relevant to my project, let me first say a few words about why many areas of psychology have no problem with replicability. The replication crisis arises, in important part, from the use of sample sizes that were too small to rule out chance as a plausible explanation of the results reported (too often, unscrupulous or surprisingly ignorant psychologists took advantage of this fact to massage data in ways more or less guaranteed to produce statistically significant results; for example, by shelving unsuccessful experiments and simply repeating them until, by chance, they got the results they wanted). But small sample sizes are only a problem in some areas. In cognitive psychology, while sample sizes (in terms of numbers of participants) are sometimes tiny, tasks are often repeated a very large number of times. The high number of trials ensures that studies have an extremely high power, and are able reliably to detect small effects. The *p* value for significance—the threshold used to assess the likelihood that evidence against the null hypothesis is due to chance—is conventionally set at 0.05. Roughly, that is, an effect is taken to be (provisionally) established, or "significant," if the probability that we would see it by chance if there were no genuine relationship between the variables of interest was 5 per cent or less. One sign that some areas of science suffer from a serious problem is that there is a suspicious clustering of published results just below the cut-off for significance (Leggett et al. 2013; Masicampo & Lalande 2012); this is evidence that researchers have engaged in *p*-hacking—manipulation—to massage the data until it reaches significance (this can be done, for instance, by dividing the data in unprincipled ways until a subpopulation is identified for whom the finding is significant). But in cognitive psychology, tiny *p* values (e.g., $p < .001$) are not uncommon (see Scholl, 2017 for discussion).

But while much of psychology is untouched by the replication crisis, I can't take a great deal of comfort in that fact. Much of the work on which I'll draw comes from social and political psychology, which are ground zero for the crisis. In the absence of better evidence, I'll draw on this work freely, albeit carefully and reflectively.

To the extent I can, I'll rely on more recent work. Methodological standards have risen dramatically since the replication crisis first came to widespread attention. Many studies are now preregistered, which

dramatically reduces the risk of *p*-hacking (if I test for a hypothesis that differs from the one I registered, or split my sample in a way that was not motivated by the hypothesis I registered, this is now plain to everyone). Sample sizes have increased dramatically, increasing the power to detect genuine effects and lowering the risk of generating a chance finding. It is now easier than before to publish null results, and even when they are not published formally, such studies are now routinely made available online. This dramatic rise in standards ensures that newer work is more likely to be reliable than older. At the same time, we now have a better sense of which older work is especially unreliable. Researchers have developed statistical tests for detecting *p*-hacking and estimating true effect sizes in the light of the file drawer effect (the shelving of experiments when they failed to cross the magic $p = 0.05$ threshold, thereby ensuring that their results do not become part of the public record), which enables us to identify unreliable work. In fact, even without formal testing researchers often have a good sense of which work is reliable and which is not (Camerer et al. 2018). If a result seems too good to be true, it probably is.

We should be cautious in drawing on cognitive science. We should be attentive to effect sizes, to sample sizes, to replicability and to how well hypotheses cohere with other work. But we shouldn't be skeptical across the board. Naturalistic synthetic philosophy—or, at any rate, my version—is motivated in part by the conviction that evidence from the special sciences is routinely better than evidence from other sources. It's not always good evidence, and I'll approach it with a skeptical eye when warranted. But on many topics, it's a far better source of evidence than philosophers' intuitions, and I'll treat it as such. I'm confident that some of the work I'll cite will prove to be flawed, sometimes seriously. But the account I'll develop should be able to withstand such blows.

For all its use of empirical findings, this is an exercise in philosophy. It depends not on the science (directly), but on interpretations of the science and on philosophical argument. The general drift of the book is increasingly philosophical: the earlier chapters describe and interpret scientific work, and the later chapters engage much more in philosophical argument. The overall conclusions run contrary to widespread views within the cognitive sciences. If the account offered here is correct, we're

much more rational than psychology and naturalistic philosophy usually holds. But our rationality depends very heavily on others and on how we're embedded in epistemic networks.

Of course, individuals are predictably overfond of their own work and predictably limited in their capacity to assess it. It is through the scrutiny of the epistemic community that my account will be tested and its strengths and flaws revealed.

1

What Should We Believe About Belief?

Beliefs, as philosophers use the term, are mental states that represent the world. We have lots and lots of beliefs, most of them very boring. I believe *that it is hot today*; *that I am in Australia*; *that it is mid-afternoon*. We tend not to notice these beliefs. We notice our beliefs more when they are controversial (*that abortion is morally permissible*; *that facemasks save lives*) and when they are identity-forming for us (*that Jesus died for our sins*; *that the US is the greatest country in the world*). The contents of these beliefs are very different from one another. A major focus of the recent philosophical literature on beliefs has been whether these different beliefs are all in fact the same kind of mental state. A belief like *the cat is on the mat* differs from the belief *the banana is unripe* in its content. The first is about cats and mats; the second about bananas and ripeness. But perhaps *the cat is on the mat* differs from the belief *Manchester United is the best team* not only in content, but also in the kind of state it is.

One reason to think that there might be different kinds of beliefs is that some beliefs are bizarre. It is hard to see how anyone could genuinely believe *that Trump is a decent person* if by "belief" we mean *take the world to be a certain way*. Moreover, beliefs have the function of guiding behavior: we represent the world as being a certain way because we need to know how to act on and in it. But some beliefs don't guide behavior. For example, *delusions* have a notoriously loose relationship with behavior. Someone who suffers from the Capgras delusion believes that some familiar individual (usually a close family member) has been replaced by an impostor. But sufferers are typically little concerned by this startling fact (Bayne & Pacherie 2005). The man who believes, or says he believes, that his wife has been replaced by a robot or an alien typically doesn't report her missing or avoid the impostor. Capgras sufferers have had a brain injury; as we'll see, healthy people are also unmotivated by some of their most fervently expressed beliefs.

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0001

I won't be tackling these issues directly here. I'm not especially concerned with the nature or structure of belief. I care about beliefs at all because of the role they play in explaining and causing behavior, and there's no real doubt that most of the beliefs that people profess are apt to cause significant behavior (even if there are interesting differences between them in this regard). A state can qualify as a belief, on the relaxed standard I'm using here, even if it falls well short of what Stephen Stich (1978) calls *inferential promiscuity* (that is, aptness to enter into inferential relations with an indefinitely broad range of propositions) or broad guidance of behavior, just so long as it drives a sufficient amount of our sufficiently consequential behavior.

In this chapter, I'll survey some of the rich philosophical and psychological literature on how beliefs are acquired and updated. I survey this literature not only for its own sake, but to set out the background for my own account. I aim to show that the existing literature, illuminating though it is, doesn't adequately explain how and why we come to believe what we do and act as we do. In setting it out, I'll also be setting out the standard and highly influential picture of ourselves that has emerged from psychology. It is now part of educated lay wisdom that human beings are far from fully rational animals, and the psychological literature on beliefs is an important part of the motivation for this piece of wisdom. I'll be arguing we're a lot more rational than the standard view holds.

Before I turn to the psychology, I'll provide some essential background on why belief matters and why we should therefore be concerned with bad beliefs. Beliefs matter because they guide behavior; accurate belief is required for successful navigation of the world. I'll also show how high the stakes are: some of the most significant political challenges of our time are, in part, battles over belief. In the second and third sections of this chapter, I'll examine accounts that hold that some of our controversial beliefs don't guide behavior in the manner characteristic of mundane beliefs; if these accounts are correct, some of our (apparent) beliefs may not matter all that much. I'll argue that we can't take comfort from these accounts: many controversial beliefs play enough of the explanatory and causal role we expect of beliefs to be extremely consequential. I'll then turn to a discussion of the family of accounts that explain bad belief as a result of

our rationality deficits. I'll give some preliminary reasons for thinking that these accounts are inadequate. The rest of the book will constitute a much fuller case for an alternative, on which we're rational animals after all.

Belief and Behavior

Belief *matters*. Many philosophers identify beliefs with functional states. On this view, to believe that *p* is to possess a representation that plays a characteristic role in behavior.[1] It *explains* our behavior (why did Anya go to the fridge? She was thirsty and believed there was beer in there) and it *predicts* behavior (will Bindi take her umbrella? She will, if she believes it might rain). Even those philosophers who reject functionalism accept that beliefs at very least typically play these explanatory roles in behavior. Beliefs are the kind of things that show up in behavior (especially, but not only, verbal behavior: assertion).[2] Attention to behavior therefore enables us to attribute beliefs. In a popular metaphor, beliefs typically function as *maps*. They enable us to navigate the world.

The (characteristic) functional role of beliefs provides us with a heuristic for belief attribution: beliefs are those states that make best sense of agents' behavior (given plausible assumptions about their desires). This fact also brings into relief why it is so important to understand the factors that explain belief acquisition and update. *Because* beliefs play a pivotal role in behavior, accurate beliefs are essential for appropriate behavior. "Appropriate," here, may mean many things. Accurate beliefs are typically required for *adaptive* behavior. Agents with accurate beliefs

[1] The focus on behavior comes easily to philosophers, but it is worth saying a few words about it for those less steeped in the philosophical literature. Focusing on behavior doesn't entail accepting behaviorism: rather, it is a way of getting at the very internal states that behaviorism ignored (or denied). A belief, on this view, is a functional state: it is the internal state of the agent that makes best sense of her behavior, given her desires. This "interpretivist" picture of belief attribution should be familiar upon reflection: if your friend says she is liberal, but votes for Donald Trump and believes that taxes should be lower, you're probably going to conclude she's lying (or self-deceived). Attributing the belief she claims to have doesn't make sense of her behavior. Some other functional state must underlie it, not the belief she professes.

[2] Rose, Buckwalter, & Turri (2014) present experimental evidence that ordinary people take assertion to be the single most powerful cue for belief ascription.

tend to do very much better at setting and achieving goals than those with inaccurate beliefs. There may be exceptions: more than a century ago, William James observed that sometimes having an inflated confidence in our abilities may help us achieve our goals. Work on "depressive realism" (Ackermann & DeRubeis 1991; Moore & Fresco 2012) also suggests that overconfidence is normal and adaptive (though the reliability of this research is now in doubt). But it's uncontroversial that accurate beliefs are generally required for effective goal pursuit. The man who succeeds in leaping across a gulf only because he has an inflated assessment of his own athletic prowess had better retain an accurate assessment of where the edge is, lest he step off it inadvertently.

Accurate beliefs are also required for pro-social behavior. Again, there are surely exceptions. Perhaps a belief in free will is false, but nevertheless required for the maintenance of social order (as Vohs & Schooler (2008) suggest).[3] Again, though, accurate beliefs are generally required for pro-social behavior. We need to have accurate beliefs about who is sentient if we are to avoid hurting others inadvertently; about what benefits them and what harms them; about where they are and where they will be, and so on. Driving a car, closing a door, simply walking across a room: all of these are actions that could cause harm, but usually do not, because we have roughly accurate beliefs about a host of banal topics.

Beliefs, and especially bad beliefs, matter not only for our banal behaviors; they're also at the heart of our major political challenges. Those who seek to manipulate our behavior have been quick to grasp this fact. They aim to manipulate us by targeting our beliefs, rather than by getting us to act directly.

In the wake of recent scandals involving Cambridge Analytica and fake news on Facebook, journalists and ordinary people sometimes express a worry that ordinary people can be easily manipulated into believing more or less anything. Actually, shifting opinions is often quite difficult: when people have pre-existing beliefs, when the topic is one that they take to concern them directly, and when they get reliable confirming

[3] This paper has been very influential. It has been cited 970 times, according to Google Scholar. Its influence indicates the difficulty of correcting false beliefs: later work finds little to no effect of free will belief on prosocial behavior (Crone & Levy 2019; Monroe et al. 2017; Nadelhoffer et al. 2020).

or disconfirming feedback, changing minds is hard and effects tend to be small and often short-lived (Mercier 2020). But when they have few pre-existing beliefs or when the topic is one on which they are not able to access direct evidence, manipulation can succeed. Unsurprisingly, we're more vigilant and careful concerning questions that touch us directly than those which seem distant and of little significance to how our lives go. The selectivity of epistemic vigilance leaves us vulnerable to manipulation on important questions when they don't make a direct difference to our lives, or we can't see how they make a difference. The obvious example here is climate change. On this topic, manipulation of public opinion has been spectacularly successful.

Most or all of my readers live in democracies, however imperfect. Democracies are responsive to the priorities of voters. Decades ago, fossil fuel interests took action to forestall voters agitating strongly for effective action on climate change. In *Merchants of Doubt* (2011), Oreskes and Conway document how they went about it, and how successful they were.

Oreskes and Conway's title stems from a now infamous 1969 memo by a tobacco executive. Big tobacco fought a long and effective battle against governmental regulation and bad publicity by casting doubt on the science linking smoking and cancer. "Doubt is our product," the unnamed author of the memo wrote, "since it is the best means of competing with the 'body of fact' that exists in the mind of the general public." The tactics the tobacco industry pioneered to engineer doubt (often, in fact, promoted by the very same individuals) came to be adopted by fossil fuel interests. The merchants of doubt don't aim to convince the public that climate change isn't genuine or isn't a serious challenge. Rather, they aim to convince us that the science isn't settled: that there is ongoing debate about these claims and that there are reasonable views on both sides. They aim in their own words, to convince the public that "significant uncertainties exist in climate science;" sufficient to cast doubt on the central claims of the IPCC (Readfearn 2015). Success for them consists in putting sufficient doubt in people's minds to ensure that calls to action are not seen as priorities, rather than in bringing people to reject the science. The tactic has been successful in bringing about higher rates of skepticism about the need for action (Capstick et al.

2015). Given the probable effects of climate change on the world, their success may be the biggest tragedy of the past half century.

The tactic of sowing doubt (rather than trying to convince people) has been adopted by campaign after campaign: the battle against the regulation of DDT, of chlorofluorocarbons, of the emissions that cause acid rain, as well as tobacco and fossil fuels. In its most recent manifestation, the tactic has taken on an even more cynical guise. The aim now is no longer to bring people to think that there may be legitimate debate about an issue, but to create confusion. Disorientation, rather than doubt, is now the product.

Allegedly, this tactic is widely used in Putin's Russia, where it is aimed at Russians themselves by state-linked agencies. Russian state media put forward a dizzying array of incompatible explanations for the downing of Malaysia Airlines Flight 17, for example, with no attempt made to reconcile them. Flight 17 was brought down by the Americans to frame Russia, *and* it was brought down by accident in a failed attack on Vladimir Putin's private jet, *and* it was brought down by Russian forces who knew that it was not a scheduled passenger plane at all. Similarly, Russian-linked bots on Twitter have not supported a particular line on vaccines; rather, they have played both sides (Broniatowski et al. 2018). They don't aim to convince viewers of some state-sanctioned narrative, "but rather to leave them confused, paranoid, and passive—living in a Kremlin-controlled virtual reality that can no longer be mediated or debated by any appeal to 'truth'" (Pomerantsev 2014). In their book-length study of the right-wing media, Benkler, Faris, & Roberts (2018) argue that the promotion of bizarre conspiracy theories by Alex Jones and the like should be understood in the same kind of way. It's not because they seriously entertain the thought that the Sandy Hook shooting was a false flag operation that the right-wing media ecosystem has devoted attention to the claim; rather, the aim (or the explanation; people need not always explicitly aim at an end to effectively pursue it) is disorientation. They aim "simply to create a profound disorientation and disconnect from any sense that there is anyone who actually 'knows the truth.' Left with nothing but this anomic disorientation, audiences can no longer tell truth from fiction, even if they want to" (Benkler et al. 2018: 37).

Should We Believe in Belief?

The apparent success of merchants of doubt in pursuing their goals by targeting our beliefs suggests that beliefs matter: what we believe explains how we act. Perhaps there's another explanation of their success, however; one that doesn't turn on belief. In this section, I will discuss accounts (due to the cognitive scientist Hugo Mercier and to the philosopher Neil Van Leeuwen), which suggest, or could be taken to suggest, that apparent beliefs matter much less than we tend to think. If they're right, there must be some other explanation of the success of the merchants of doubt.

Mercier, Van Leeuwen, and others who have pursued this kind of line draw our attention to an important fact. They're right that people do not (straightforwardly) believe many of the things they profess to believe. Still, I'll argue, differences in our beliefs (in states that are beliefy enough to count for our purposes) are at the heart of many of the central political issues of our day, such as disputes over climate change and vaccination. What people say they believe tends to explain and predict how they will act. Let's begin with the case for thinking that many of the beliefs people express are relatively inert or epiphenomenal.

Reports of startling ignorance or of bizarre beliefs are staple fare for the media. Consider some recent headlines:

One-third of Americans think alien UFOs have visited Earth (Whalen 2019).

One-third of Americans don't believe 6 million Jews were murdered during the Holocaust (D. Brennan 2018).

One in four Americans think Obama may be the antichrist, survey says (P. Harris 2013).

Nearly half of Republicans see truth in "Pizzagate" theory: Poll (Blake 2016).

Polls like these suggest that a large proportion of our fellow citizens are fundamentally disconnected from reality.

Consider the "Pizzagate" conspiracy mentioned in the last headline for an egregious example. In March 2016, John Podesta, the chairman of Hilary Clinton's election campaign, had his email hacked. Several months later, a number of these emails were published by Wikileaks. Apparently as a joke (Kunzru 2020), some people posting on social media and discussion boards began to suggest that the emails contained code words referring to pedophilia, and that senior figures in the Democratic party were involved in the trafficking of children for sex. Allegedly, food related words in the emails referred to sex with children (for instance, "cheese pizza" meant "child pornography," since they share initial letters).

As the conspiracy theory transmuted, it came to be associated with the Comet Ping Pong restaurant, a Washington D.C. pizzeria that had hosted fundraising events for Obama. Comet Pizza was alleged to host parties involving sexual abuse of children. At least some of the people who fabricated ever wilder versions of these stories or who helped spread them were trolls or pranksters. Some, in fact, probably thought that the stories were harmless jokes because they were so implausible. But others took them seriously. On December 4, 2016, Edgar Maddison Welch drove to Comet Pizza with an assault rifle. He was there, he said, to investigate the claims and free any child slaves he found. He fired several shots before he was arrested.

As Pizzagate illustrates, not everyone who promotes bizarre conspiracy theories genuinely believes them. In the face of the sheer wildness of Pizzagate and some other conspiracy theories (the moon landings were fake; the government is hiding the bodies of aliens who crashed at Area 51 in Nevada; the British Royal family is heavily involved in trafficking heroin, and so on), we may wonder whether the Welches of this world, those who take these theories at face value, are in fact a tiny minority. Perhaps most of the people who report belief in these conspiracies are pranksters or trolls, or aiming to fit in with others who they take to believe them. Perhaps belief reports are explained by something like the phenomenon called *pluralistic ignorance*, whereby most or all people conform to a norm they don't themselves take to be justified because they think a majority of others take it to be justified (D. T. Miller & McFarland 1991).

Psychologists—if not pollsters—have long recognized that it can be very difficult to probe people's *real* beliefs. On many topics, we lack

settled beliefs: asked what we believe, we may take a side rather than admit ignorance, because admitting ignorance can be threatening to our self-image (see Motta et al. 2019 for evidence that giving people the option of skipping a question—giving them a "soft don't know" option—has a significant impact on responses compared to requiring them to *report* they don't know). Many responses are produced simply for the purpose of answering the question asked and not to report a pre-existing mental state. We therefore ought to take reported support for conspiracy theories and other bizarre beliefs with several grains of salt. Perhaps people are more rational than these surveys suggest, or at any rate have more realistic beliefs.

There are a variety of reasons why people may not report what they genuinely believe: they may be deceptive in their reports, or they might be engaging in some kind of social signaling, rather than reporting (Funkhouser forthcoming; Mercier 2020; Sterelny 2015, 2018). Equally, they may not really believe what they sincerely report: they may be mistaken about what their beliefs are. These are all possibilities that I take very seriously: in fact, I think each of them explains some belief reports. Even taken together, however, they don't give us reason to think that our exemplars of bad belief are typically not genuine beliefs (that is, beliefy enough to guide a wide class of significant behavior).

Let's begin with the cognitive scientist Hugo Mercier. Following Sperber (Sperber 1997), Mercier (2020) distinguishes between intuitive and reflective beliefs; only the former govern behavior broadly, he suggests. He argues that we are liable to (reflectively) accept bizarre conspiracy theories and rumors *because* they do not have "serious practical consequences" (159) for our behavior. One of his primary examples is the rumor that swept the town of Orleans in 1969, that Jewish shopkeepers were kidnapping local girls and selling them into slavery. Mercier notes that people didn't take the kinds of steps that full belief in such stories would seem to warrant (demanding police action or raiding the shops, for example). Like the Pizzagate conspiracy theory, for most people these assertions were causally insipid: they motivated only low cost and relatively trivial kinds of behavior (Mercier reports that at the height of the rumor, some people stopped and stared at the shops alleged to be involved; hardly a proportionate response to slave trading). Perhaps

there are odd examples of people who go on to act consistently with their theories, but their behavior is probably best explained by some quirk peculiar to them, not the apparent belief they share with many others.

Why do people profess bizarre beliefs, if these states don't govern very much of their behavior? Mercier gives three reasons. First, professions of belief may function as a kind of commitment device. Because these beliefs are bizarre and at odds with the respectable consensus, professing to hold them serves as a reliable signal of allegiance to a group. Talk is usually cheap, but willingness to engage in talk that paints one as a kook is not cheap and is therefore a good signal of group belonging. Second, people may be willing to accept certain claims (say, "that Obama is a secret Muslim") because these beliefs have little effect on their behavior: when we don't expect to be able to act on an assertion, we have little incentive to test its plausibility. Third, people accept and repeat rumors to justify how they wanted to act in any case, rather than to acquire new information that is consequential for further behavior. Voltaire is often paraphrased as saying that "those who can make you believe absurdities can make you commit atrocities." Mercier claims this is backwards: "it is wanting to commit atrocities that makes you believe absurdities" (202).

Mercier gives several striking illustrations to back up his claims of the relative inertness of the beliefs people express. One comes from the history of medicine. On a naïve view, he suggests, the millennia-long persistence of bloodletting in the Western world as a medical treatment could be attributed to the influence of the ancient Greek Hippocratic writers and those who developed their ideas, in particular the hugely influential Galen. But that would be wrong, Mercier claims: bloodletting is practiced in at least one fourth of the world's cultures, including cultures very far removed from Greek influence. It is practiced in small-scale societies in Central and South America, in Africa, and in South East Asia. This indicates that it was because bloodletting is intuitive, not because of the influence of the Galenic manuscripts, that it was practiced. Another illustration Mercier provides is the blood libel (i.e., that Jews kill Christian children for ritual purposes). The blood libel was a common accompaniment of pogroms, but these rumors sometimes spread without precipitating violence and pogroms were sometimes accompanied by much milder rumors. Both examples suggest that people look to

apparently factual claims to justify their behavior, rather than using them as a map to steer by.

But Mercier's examples don't show that these professed beliefs were causally inert or insipid. That many societies with no exposure to the Galenic manuscripts practice bloodletting shows that the influence of Hippocratic medicine wasn't necessary for such practices to develop. It doesn't show that bloodletting would have been practiced in medieval Europe in the absence of Galenic influence. After all, most societies *don't* practice bloodletting. Moreover, even if the Galenic influence wasn't necessary for the development of bloodletting in Europe, that influence *wasn't* causally inert, on Mercier's own telling: bloodletting in Europe had distinctive features (for example, drawing much more blood than seen elsewhere) due to the influence of Galen.

Mercier is right to emphasize that beliefs take hold only under certain conditions. The inhabitants of Orleans, or of Kishinev where an infamous pogrom resulted in the death of 49 Jews in 1903, accepted outlandish stories because they were already disposed to feel ill-will toward Jews. They embraced a handy justification for their ill-will and (in the latter case) of the atrocities they went on to commit. But the justifications don't seem to have been inert: without them, the acts may not have taken place, or may have been less widespread or less serious. For some people they functioned as an excuse, but for those on the fringes they may have functioned as a reason. In fact, it seems impossible to explain the events at Orleans or Kishinev except by citing the kinds of rumors and misinformation Mercier claims to be inconsequential. We can't explain these events by citing envy or worries about competition alone, because that explanation leaves certain important facts mysterious. Why the Jews and not other successful shopkeepers, or government officials? Handy scapegoats are handy because they are stigmatized, and stigmatization involves belief. The rumors may have taken hold only because people were already disposed to hate and despise Jews, but they were disposed to hate and despise Jews in very important part due to centuries of previous rumors and propaganda.

At the same time, Mercier is persuasive that it is much harder to shift people's firm beliefs than is usually thought. Propaganda, advertising, and rumor have very limited power to move people to reject a belief that

is entrenched to any significant degree (on the other hand, as we will see, it is often trivially easy to shift people from one belief to an opposed belief when the first is *not* entrenched, even when people take themselves to be fervently committed to it). He is also persuasive that we are very much less likely to accept rumors and conspiracy theories when they are seen by us as having "serious practical consequences" (159) for our lives (as opposed to the lives of others for whom we have little sympathy).[4] But the limited power of propagandists is sometimes enough to bring about serious consequences.

Propagandists can play to our prejudices and make our beliefs more extreme and us more likely to act on them. They can also sometimes take advantage of decreased vigilance with regard to beliefs we see as inconsequential for ourselves by dissimulating the consequences, and they can have spectacular success when the consequences are distant and abstract. Were the consequences of our beliefs about climate change more easily perceptible and more personal, we might be less apt to accept conspiracy theories about it. Because the consequences are far removed from individual behavior, however, we aren't vigilant. Our relative credulity on this topic helps to explain why we face a climate crisis with little political will to address it.

Neil Van Leeuwen's (Van Leeuwen 2014, 2018, forthcoming) focus is especially on religious beliefs. He argues that religious beliefs are not factual beliefs: they belong to a different class of representations, which play a different role in behavior. They play an identity-constituting role; any role in guiding behavior is more constrained and secondary to this primary role. His case is built around evidence from cognitive science, which seems to indicate that religious beliefs don't consistently guide behavior. Van Leeuwen maintains that ideological credences typically belong to the same class of representation as religious beliefs; they're identity-constituting, not action-guiding. His view is therefore a

[4] The COVID-19 pandemic has provided some telling examples of how people may come to accept conspiracy theories with tragic consequences. While in most Western countries, these conspiracy theories have (at the time of writing) caused little consequential behavior beyond the apparent targeting of 5G mobile phone towers, the rumor that alcohol can prevent or cure illness has led to hundreds of deaths in Iran (Associated Press 2020). Sometimes, fake news kills.

challenge to one (like mine) that holds that bad beliefs matter because of the relatively broad role they play in behavior.[5]

Evidence of mismatches between professed religious beliefs and behavior is plentiful.[6] Christians, Hindus, and Muslims seem more likely to behave consistently with the tenets of their religion when it's salient to them than at other times. For example, Christians are more charitable on Sundays than on other days of the week (Malhotra 2010); they also appear to consume less pornography on Sundays (in highly religious US states, consumption of internet porn is lower than in less religious states only on Sundays (Edelman 2009)); Malhotra dubs this propensity of Christians to behave more consistently with their professed beliefs on one day of the week the *Sunday Effect*. Analogous effects have been reported for other religions. Xygalatas (2012) found that Hindus who played an economic game in the context of a temple withdrew significantly less money from a pot that would otherwise benefit the entire group of players than did those who played the game in a secular setting. Similarly, Duhaime (2015) found that Muslim shopkeepers in Morocco gave significantly more to charity within 20 minutes of hearing the call to prayer than at other times. Of course, the faithful claim to accept ethical norms that govern behavior at all times, not just in certain settings, but their behavior suggests that these norms have a much weaker grip on them outside contexts that remind them of their commitments.

Other instances of apparent mismatch in the religious domain don't involve pro-social norms. Consider the phenomenon that has come to

[5] Van Leeuwen isn't committed to denying that bad beliefs explain a lot of consequential behavior. He accepts that there may be some ideological beliefs (and some religious beliefs) that are held as factual beliefs, though he thinks they are atypical, at least today. He also accepts that the "secondary attitudes" govern *some* behavior, and that might be enough to make them troublesome. Given that (as we shall see) he predicts that secondary attitudes tend not to guide behavior in high-stakes situations (except when the stake is the agent's identity), though, he seems committed to holding that secondary attitudes are less directly and significantly troublesome than I take bad beliefs to be.

[6] Alas, plentiful data is not always reliable data. The sample sizes for most of this work are small and little of it was preregistered. I have more confidence in the data underlying the Sunday effect (if not necessarily the explanation offered) than in the other reported results. It is intrinsically difficult to gather large samples for experimental work that aims for ecological validity. The Sunday effect is based on correlational data, which is part of the reason the sample is much bigger.

be called *theological incorrectness*. Theological incorrectness occurs when people who profess to believe the official tenets of their religion appear to utilize conflicting representations in interpreting religious stimuli; implicitly attributing limited knowledge or a limited capacity for attending to events to God, for example (Barrett 1999). Similarly, there is evidence that afterlife beliefs are context dependent. Harris & Giménez (2005) and Astuti & Harris (2008) found that Spanish children and Vezo children and adults (respectively) were more likely to attribute continuing mental life to the recently dead when the person was described as dying in a context that featured primes for religion, such as an attending priest, rather than in a more secular context.

Mismatches like these suggest that people don't factually believe the tenets of their religion. Factual beliefs predict and make sense of our behavior across all contexts to which they're relevant; religious beliefs appear to lack this property. People act consistently with their religious beliefs only when they're salient to them. Van Leeuwen takes this functional difference to be good evidence that they belong to a different class to factual beliefs.

According to him, religious beliefs are closely akin to (perhaps even identical to) imaginings. Imagination, too, only guides our behavior in certain contexts (only when Wendy is playing *fire engines* does she cover her ears in response to the noise of the siren she imagines; outside the game, her truck is silent), whereas belief guides our behavior all the time (even while she's playing fire engines, Wendy doesn't worry that her tree house will catch fire, and she still takes care to stay away from the edge). Imagination requires effort to be sustained: imagined states tend to fade quickly and reality—ordinary belief—takes over. The anthropologist Tanya Luhrmann (2012) has proposed a similar account. She argues that the relationship to God her informants—members of the Vineyard Evangelical Church—cultivate is half recognized by them as an effortfully sustained imagining.

I'm deeply skeptical of much of the evidence for the claim that religious beliefs function differently to factual beliefs. Much of it comes from the (too recent) bad old days of psychology, and reports surprisingly large effects that seem more likely due to chance than to the detection of an underlying reality. Nevertheless, I'll proceed under the assumption that

the data is reliable. Even on that assumption, we should still think that religious beliefs are (typically) beliefy enough to count for our purposes: they still guide a great deal of consequential behavior.

Van Leeuwen argues that religious beliefs tend not to guide highly consequential behavior, because religious "believers" operate with what he calls a "two-map cognitive structure" (Van Leeuwen 2018, forthcoming). One map represents the world as it is factually taken to be; the other map represents the world as they religiously represent it. The faithful never lose sight of the real world in their religious imaginings. They (implicitly or explicitly) track the difference between the world as they religiously imagine it and the way the world really is (just as Wendy tracks her tree house, even while she's playing fire engines). They are thereby able to monitor the gap between the two maps. This monitoring enables them to steer away from putting their religious beliefs to too rigorous a test. For example, believers sometimes call on God to perform miracles, but they're careful to ensure that they ask only for things that might happen anyway. They pray for rain, not showers of money; they pray that someone's cancer goes into remission, but not that a leg grows back (Barrett 2001). This two-map cognitive structure also ensures that religious believers avoid paying the high costs that might be associated with factually believing the tenets of their religion. When stakes rise, Van Leeuwen claims, factual reality rushes back. The devotee may claim that God looks out for her, but continue to take out health insurance.

But believers often *do* engage in behavior that seems to make sense only if they really—factually—believe much of what they claim. They make large donations to churches. They pay for masses to be said for them after their deaths (when they can no longer reap any ancillary benefits). They refuse conventional health care in favor of an exorcism or a faithhealer. They refuse lifesaving blood transfusions. They may do so on behalf of their children as well as themselves. Medieval trials by ordeal provide a dramatic illustration of how religious beliefs may shape behavior in high-stakes contexts.

Trial by ordeal was usually used when evidence was sparse or contradictory. It gave the accused the opportunity to clear their name by performing a painful ritual. Trial by fire (walking a set distance over red-hot

ploughshares or clutching a red-hot iron) or trial by water (plunging one's hand into boiling water to retrieve a stone) were the most common ordeals. Trial by ordeal was held to be determinative of guilt on the basis that God would protect the innocent, by preventing injury or by speeding the healing process. To contemporary eyes, trial by ordeal seems not only barbaric but useless: we'd expect everyone who underwent it to be found guilty. The available evidence tells a different story: a surprisingly high proportion of those who underwent trial by ordeal were acquitted.

Peter Leeson (2012) points out that the rituals surrounding trial by ordeal gave the attending priests ample opportunities to manipulate the results. They could, for example, exert a lot of control over the temperature of the irons. Why would they engage in such manipulation? Leeson suggests that manipulation might have occurred in response to the demeanor of the accused. Those who faced the trial with relative equanimity thereby gave evidence that they believed God would protect them, and unintentionally signaled innocence to the priests. Conversely, those who were more fearful thereby provided evidence of their guilt and suffered accordingly. The contemporary evidence indicates that such manipulation must have been selective: faced with trial by ordeal, many people pleaded guilty rather than undergo it (and face the harsher penalties associated with being found guilty in this way to boot).

If the behavior of those who faced trial by ordeal really provided evidence in this kind of way, however, then they must have really—factually—believed that their innocence was protective. There's no sign here of a two-map cognitive structure: rather their religious beliefs seem to function as factual beliefs and govern behavior, in this very high-stakes context, accordingly. Of course, this leaves the behavior of the priests unexplained. If they factually believe that God will protect the innocent, why do they intervene? Leeson suggests that the priests might come to see themselves as serving as God's instruments in manipulating the trials in this kind of way.

Van Leeuwen always acknowledged that some religious beliefs might be held as factual beliefs. In his most recent work he goes further, suggesting that religious credences, as he calls them, may not be seen in all cultures or all religious communities. Perhaps some communities

accept most or all their religious beliefs factually.[7] He claims that religious credences are typical today, not universal. Especially given this restriction on the scope of his view, it's difficult to produce decisive evidence against it. There are many reasons why people may behave inconsistently with their factual beliefs, after all, and many why they might behave consistently with their imaginings. Our implicit attitudes may conflict with our genuine beliefs, and these attitudes sometimes cause behavior (Levy 2014a, 2015). We may have inconsistent beliefs (Brownstein et al. 2019; Mandelbaum 2016). We may have in-between beliefs that conflict with our genuine beliefs (Schwitzgebel 2010). We may be mistaken about what we believe, because we haven't reflected deeply, and because our beliefs are not open to introspection (Carruthers 2013). Motivated cognition (Kunda 1990) may make our beliefs insensitive to evidence and to the context we find ourselves in. All these phenomena are consistent with Van Leeuwen's account and could explain divergences between the behavior he predicts and what we actually observe.

This dizzying variety of alternative explanations of departures from behavior consistent with belief and with imaginings renders a search for a decisive test of Van Leeuwen's account futile. What matters for our purposes is whether our paradigms of bad belief (concerning climate

[7] In his new book, Van Leeuwen emphasizes ethnographic work on the Vineyard movement, an evangelical Christian church that began in the United States and has since spread to many other (developed) countries. To what extent the Vineyard movement is representative or typical of religion worldwide is, however, a difficult question. Members of the church live in secular societies, in which science has a great deal of prestige and in which many scientific claims are more or less universally known. Perhaps the Christian who lives in the United States doesn't take a factual attitude to the proposition that *God created the Earth in 6 days less than 10,000 years ago*, given she knows that this proposition is rejected by scientists, and she knows that science is widely held to be the most reliable way of discovering the truth of propositions like this. But Christians who asserted that same proposition in medieval Europe were in nothing like that position. Theistic creation may have seemed (perhaps *been*) the most plausible explanation of the origin of the world available. That medieval Christians took a factual attitude to propositions like this one seems to be overwhelmingly plausible; so, perhaps, do religious believers in other parts of the world, in which there is less access to good education and/or science is held in lower regard. A more difficult question is whether people at such times and in such places took and take a factual attitude to propositions like *there are ghosts and other spirits who cause good and bad events in the world*. They have and had access to naturalistic explanations of (many) such events, which might have militated against such beliefs. But given the lower prestige of such explanations in these times and places, they may have given them less credence and been more inclined to factual belief in non-naturalistic propositions.

change or vaccinations, for instance) are in fact sufficiently belief-like to help explain and predict a great deal of consequential behavior. In the original paper introducing his theory, Van Leeuwen (2014) suggested that those who reject the science of climate change might hold an essentially religious attitude toward certain factual propositions. He's right there are good reasons to think that climate change skeptics often don't have very determinate beliefs. They seem to oscillate between believing *that climate change isn't happening*, *that it's happening but we're not causing it* and *it's happening and we're causing it, but it'd be too expensive to fix*, depending on which is handiest. The fact that they move between incompatible propositions suggests that they don't have a very determinate or stable belief (they're not unusual in that; Levy 2018). Nonetheless, their behavior is best explained by something beliefy. Climate change skeptics aren't merely ignorant: they don't just fail to know that climate science is true. Most qualify as skeptics at least in part in virtue of holding a distinctive belief; and it's a belief with a fairly precise content, even if it shades into imprecision when they attempt to flesh it out. The settled content is the content that's common to all the inconsistent propositions they oscillate between. They believe something along the lines of *climate change isn't a problem we need to address.*

Unless we attribute to climate denialists a belief with a content along those lines, we can't begin to explain their behavior. This belief makes best sense of their verbal behavior (they *say* they reject the science of climate change), and what people say is a good, though far from infallible, guide to what they believe. But more than that, this belief makes best sense of their voting behavior, their consumption habits, and the ways in which they invest their time and energies. If you have a stake in the future—if you're young enough to expect to live to see the dramatic changes that climate change will bring even to wealthy individuals in wealthy countries, or if you have children—then failures to support action to rapidly reduce emissions and seek to induce other nations to do the same is good evidence that you don't actually accept the science.[8]

[8] It's not conclusive evidence, of course. If you are completely fatalistic about climate change, you might accept the science but lack any motivation to address it. So called "doomers" are a real phenomenon, but doomers don't exhibit the combination of attitudes seen with skeptics: verbal denial of the science of climate change (remember, "denial" here encompasses saying that you believe it's real but not a priority) together with lack of motivation to do anything about it.

The science very clearly entails that significant action is urgent, if we and our children are to have decent lives (IPCC 2018). Beliefs motivate behavior in conjunction with desires: given that we can safely attribute the desire to maintain a habitable planet and a comfortable standard of living for their children to the majority of skeptics, we can only explain their behavior by attributing genuine—factual—disbelief to them, and therefore a belief *that there isn't a problem we need to address*.

We'll return to the apparent inconsistency in behavior exhibited by religious believers (and plenty of other people) in a later chapter. Let's turn, now, from the domain of religion to that of ideology, and examine evidence that partisans may not believe some or much of what they assert.

Expressive Responding

For the past few years, a favorite sport of many US liberals has been mocking the credulousness of the supporters of Donald Trump. The single most shared election-related story on Facebook in the three months leading up to the 2016 election was literally unbelievable: *Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement* (O'Connor & Weatherall 2019). But that's far from the most bizarre thing that Republicans have apparently believed. In one survey, 20 percent of Republicans reported believing that Obama is the Antichrist (P. Harris 2013). Consumption and sharing of fake news online is disproportionately common on the political right (Benkler et al. 2018), but the left is by no means immune. Fake stories congenial to the right were shared around 30 million times during the three months before the 2016 election, and those congenial to the left were shared around 8 million times (Allcott & Gentzkow 2017). Fake news appears to spread further and faster than real news (Vosoughi et al. 2018).

How could anyone possibly be taken in by these often-bizarre claims? There is some evidence that older people are considerably more likely to share fake news than younger: one study of online behavior during the 2016 election found that Republicans over 65 were seven times more likely to share fake news on Facebook than people of any political

leaning aged 18-29. Overall, around 11 percent of over 65s shared fake news (Guess et al. 2019). One possible explanation is that this group is less internet savvy or more trusting (having grown up in an age with fewer and arguably more trustworthy media outlets) than younger people, and therefore more easily taken in. But there are other plausible explanation for the sharing of fake news and the expression of support for partisan falsehoods, including some that suggest that people aren't all that credulous after all.

Around a third of Americans report believing the "Birther" conspiracy theory, according to which Barack Obama was not born in the United States (Uscinski & Parent 2014). However, there's good reason to think that the surveys overestimate the true extent of belief. On these highly politicized questions, agents may engage in *expressive responding* (Berinsky 2018; Bullock et al. 2015). Expressive responding occurs when people report beliefs to express their support for their "side," rather than because they genuinely hold them. Expressive responding may help explain some mismatches between people's reported beliefs and their behavior. Surveys have long documented large partisan gaps in attitudes, with each side perceiving the world in a way that seems to conform to their normative views (Lerman, Sadin, & Trachtman 2017). Republicans and Democrats report divergent beliefs about factual claims (e.g., the effects of economic policies). But there is some—albeit mixed—evidence that they don't go on to act in ways that are consistent with these professed beliefs (see Bullock & Lenz, 2019 for review).

There's persuasive experimental evidence that people sometimes engage in expressive responding. Schaffner and Luks (2018) took advantage of the controversy over Trump's inauguration crowds to probe its extent. The Trump administration notoriously claimed that the crowd was the largest ever to witness an inauguration, claims that flew in the face of the photographic evidence (it was this incident that led Kellyanne Conway to introduce the phrase "alternative facts" to describe administration claims). On the two days immediately following the controversy, Schaffner and Luks showed participants photographs of the Trump and the Obama inaugurations (without identifying them), and asked which depicted a larger crowd. Given how widely reported the story was, they

knew that many participants would recognize the photos and be aware of their sources. A very small proportion of non-voters and Clinton voters identified the photo of the Trump inauguration as depicting a larger crowd (3 and 2 percent respectively). In contrast, 15 percent Trump voters identified the image of his inauguration as depicting the larger crowd. It's hard to believe they were reporting a genuine belief: the photographic evidence was clear. Instead, it seems that many people are willing to report a belief they don't hold in order to express support for their preferred party or candidate. Schaffner and Luks note that some Trump supporters were probably unaware of the controversy or failed to recognize the photos, and therefore didn't see the task as presenting them with an opportunity for expressive responding, so it may be that the percentage of people willing to respond expressively is higher than 15 percent.[9]

Other studies have used different methodologies in an attempt to measure the prevalence of expressive responding. Incentives seem to be effective in reducing the partisan gap in responses. For instance, Prior et al. (2015) found (relatively) small monetary rewards for correct responses halved partisan bias (from 12 percent to 6 percent). Bullock et al. (2015) report similar results, and an apparent dose-dependence of reduction: the larger the incentive, the bigger the reduction in bias. A combination of treatments apparently succeeded in eliminating bias altogether. In contrast, Berinsky (2018) found little or no evidence of expressive responding on the questions he probed, despite offering an incentive (albeit one of a different type: a reduction in time spent on the survey). Taken together, and despite some failures to narrow the partisan gap via the provision of incentives, the evidence suggests that a substantial number of survey respondents knowingly and deliberately misrepresent their true beliefs for expressive purposes.

[9] As Michael Brownstein pointed out to me, this suggestion is supported by the fact that better educated Trump voters were much more likely to say the smaller crowd was bigger, presumably because they were more likely to recognize the photo. When Schaffner and Luks told participants of the source of the photos (in the second experiment), the effect of education vanished.

In fact, the studies to date may underestimate the extent to which people fail to report their prior true beliefs. First, incentives for accuracy might have perverse effects: they provide an opportunity for more powerful expressive responsive. The bigger the monetary reward forgone, the stronger the signal a belief report sends. For this reason, we should expect the most partisan participants to be difficult to shift by monetary reward. Further, if participants count support for a person, policy, or stance as a sacred value, they'll likely spurn the opportunity for financial reward for accuracy: sacred values are usually held to be incommensurable with and tainted by financial reward (Tetlock 2003).

A second reason why incentivization may not result in people reporting their true beliefs is that people may often lack any prior belief at all. Political scientists have long recognized that a substantial proportion of survey respondents construct their responses on the spot (J. Zaller & Feldman 1992). Instead of responding expressively, some participants may use partisan heuristics, biased sampling methods or motivated inference to generate a response, in the absence of a prior belief. To the extent to which this occurs, surveys of public opinion play a role in producing the responses they aim to probe. Someone who reports believing *that Obama is a secret Muslim* or *that Hilary Clinton gave uranium to Russia in exchange for donations* may not believe these things prior to being asked. Rather, they engage in biased memory search or biased inference procedures, or apply heuristics, to construct a belief report. In part, this may be explained by an aversion to admitting ignorance: Bullock et al. (2015) found that while only 15 percent of respondents gave a "don't know" response in the absence of incentives for accuracy, provision of incentives dramatically increased selection of the option. While the responses given *may* persist as beliefs for some time after the survey, surveys that construct such attitudes overreport their prevalence in the population.

Taking expressive responding and overreporting as a consequence of belief (or attitude) construction into account, there's little doubt that surveys often exaggerate the extent to which people genuinely believe fake news and the like. Still, there's also little doubt that substantial numbers of people do accept some fake news some of the time sufficiently to have an influence on their consequential behavior. In 2017 for example, a

false story that the founder of Ethereum had died in a car accident caused the market value of the company to drop by $4 billion (Dunning 2019). More disturbingly, a number of conspiracy theorists have escalated their harassment of parents who lost children in school shootings beyond online trolling to confronting and threatening them in person (Raphelson 2018; Robles 2016). Similarly, people who deny the existence of Covid have invaded UK hospitals and attempted to remove patients from ventilators (Quinn & Campbell 2021). While these conspiracy theorists are, no doubt, outliers, other differences in behavior are large enough for us to be confident that they aren't driven by outliers. For instance, Lerman, Sadin, & Trachtman (2017) report that Republicans don't merely *say* that they distrust Obamacare; they are also less likely to enrol in it (see Bullock & Lenz 2019 for further examples).

It's likely that estimates of the prevalence of anti-vaxx beliefs are inflated by expressive responding and attitude construction (for the record, I predict that the COVID-19 vaccine will be taken by significant numbers of people who, on surveys, report that they won't), but assertion of anti-vaxx sentiment correlates with behavior: anti-vaxxers vaccinate their children at lower rates, leading the World Health Organization to list vaccines hesitancy as one of the top ten threats to health in 2019 (WHO 2019). It's hard to explain willingness to put one's children at risk expressively; nor are the majority of people who refuse to vaccinate those who have previously participated in a poll. Similarly, even if reported skepticism about climate change is exaggerated by expressive responding or a desire to "own the libs," the fact that those on the right tolerate and encourage inaction on climate is evidence they genuinely believe it isn't a significant problem. While expressive responding might explain a large proportion of those who espouse really bizarre claims, with only a few outliers really buying into Pizzagate or the Sandy Hook conspiracy theories, it can't explain away some of the most consequential cases.

To this point in the chapter, we've been considering evidence that bad beliefs are not a pressing problem. I've argued that the evidence suggests that reports of bad belief are exaggerated: people are less credulous than we often think. But that's little comfort: the most consequential cases continue to center, very significantly, on bad belief. Your neighbor may not believe the QAnon guff he posts on Facebook, but he genuinely

doesn't believe that climate change is a problem requiring urgent action. Bad beliefs should continue to be our focus. With that in mind, let's turn to consider theories of bad belief formation.

Deficit Accounts

On the political right, acceptance of the science of climate change is much lower than in the center or the left. While 66 percent of Democrats support policies aimed at reducing or mitigating global warming, only 27 percent of Republicans express support for such policies (Funk & Kennedy 2019). Republican skepticism is not motivated by doubts about the efficacy of these policies as a means to address the problem: it is motivated by skepticism about the *existence* of the problem. Only a minority of Republicans, and a small minority of those who identify as conservative Republicans, report believing that climate change constitutes a serious threat (B. Kennedy & Hefferon 2019). The gap between Republicans and others is likely inflated by expressive responding and the like, but it's too large to be eliminated by these considerations and there are good reasons to think that on this topic, the skepticism expressed is largely genuine. Climate change is a central, and crucially important, illustration of bad belief, but it's by no means the only one. Around half the population of the United States rejects the theory of evolution, for example (Newport 2014).

In their broad outlines, neither evolution nor climate science is remotely controversial among scientists. Not only is there a scientific consensus on the reality and the urgency of anthropogenic climate change; there is also a scientific consensus on the existence of the consensus (Cook et al. 2016). Moreover, neither are matters on which the general public is well positioned to dissent from expert opinion. It is not as though the kind of evidence on which the theory of evolution is based is easily available to the general public, or that rival accounts could easily be tested without an enormous amount of specialized knowledge and tools. Why do so many people have beliefs at odds with the scientific consensus? In the light of the evidence surveyed in *Merchants of Doubt*, some kind of disinformation program—by fossil fuel interests

in the first case, and from the pulpit in the second—seems a plausible explanation.

On this kind of account, bad beliefs might be due to an *information* deficit. People accept false claims because they haven't been exposed to better information. Information deficit accounts are popular, but *rationality* deficit accounts are perhaps even more common. On these accounts, bad beliefs are not due to bad information, but some kind problem in *processing* information. Accounts of this sort are typically motivated by work in the cognitive sciences, allegedly demonstrating widespread irrationality. A third, related, kind of account is inspired by work in philosophy rather than in the mind sciences. *Virtue* deficit accounts explain bad belief formation as arising from a lack of one or more intellectual virtues. Of course, these accounts needn't be exclusive: perhaps bad beliefs arise from a variety of causes.

In the rest of this chapter, I'll discuss the first two kinds of deficit account. I'll leave virtue deficits for a later chapter. My aim is to show that these accounts face problems sufficiently large to motivate the development of an alternative. While all three may well explain many cases of bad belief formation, even in combination they fall well short of explaining the kinds of cases I'm focusing on here. We need some alternative. Developing that alternative is the goal of the rest of the book.

(a) Information Deficits

Information deficit accounts are familiar. People often blame bad beliefs on a failing school system, or on exposure to Fox News, or on echo chambers on Facebook. Books like *Merchants of Doubt* provide evidence that supports these theories: if tampering with people's information didn't work, presumably companies like Exxon wouldn't have spent so much money doing it (despite knowing that the narrative they were promoting was false (Banerjee et al. 2015)).

Surely there are people who don't accept the science of climate change because (for whatever reason) they lack access to good information. But there's good evidence that this isn't the principal explanation. Disbelief in climate change is predicted by political allegiance, not misinformation.

Being on the political right, and especially on the pro-market right, is the single best predictor of climate change denial (Lewandowsky, Gignac, & Vaughan 2013; S. van der Linden et al. 2021; McCright & Dunlap 2011; Schuldt et al. 2011). Rejecting the scientific consensus may correlate with misinformation about one important topic: skeptics are apt to underestimate the degree of scientific consensus (Cook et al. 2016; Leiserowitz et al. 2014; S. L. van der Linden et al. 2015). But this apparent fact aside, climate change skeptics don't seem worse informed than those who accept the science.

The Ordinary Science Intelligence test measures people's basic scientific literacy. For non-politicized topics, OSI scores predict accuracy in belief. Those with higher scores are more likely to answer correctly when asked whether electrons are smaller than atoms or to identify which gas is most plentiful in the Earth's atmosphere. But the neat correlations dissolve when people are asked about politicized topics. While the positive correlation between OSI scores and accurate beliefs about climate change and evolution holds for (so-called) liberals, it fails for those on the political right and those higher in religiosity (respectively). Being well informed about the mechanisms of natural selection doesn't predict accepting the theory of evolution among those higher in religiosity (Lawson & Worsnop 1992). Equally, knowing what scientists say about climate change doesn't predict accepting the scientific consensus for those on the political right (Funk & Kennedy 2019; Kahan 2015). While there may be a correlation between lack of information and bad belief formation, those who reject the science don't do so *because* they lack information. If anything, the causal arrow probably points in the opposite direction; people are unmotivated to seek better information because they take it to be irrelevant or misleading.

Of course, evolution and climate change are unusual issues. On many other topics, we do find a strong correlation between lack of accurate information and false beliefs, or lack of any belief. That should be obvious. If my only source about the weather in Chicago is misleading, it's likely that I'll acquire false beliefs about the weather in Chicago. If pressed, I'll have to confess that I have no idea when the next train from Milan to Verona departs (or even whether there is such a train). My lack of information about the timetabling of Italian trains explains my lack of

belief here. Some common false beliefs (including many pseudoscientific claims—e.g., that people are left or right brained, or that we only use 10 percent of our brains) are surely explained by misinformation. But there's little reason to believe that lack of, or bad, information is the principal explanation of bad beliefs about highly contentious topics like climate change.

There's more evidence against information deficit accounts. But that evidence is probably best discussed in the next section, which focuses on a lack of rationality rather than of information.

(b) Rationality Deficit Accounts

Rationality deficit accounts explain bad beliefs by reference to how information is processed (see Bardon 2019 for a recent book-length defence of this kind of account). Folk psychology and scientific psychology both recognize the (apparent) existence of *motivated reasoning*. When we find a conclusion unpalatable, we engage in intellectual contortions to reject it. We seem to take mixed evidence to support our prior views (Lord et al. 1979), for example, because we engage in *biased assimilation*, applying more lenient standards to evidence we take to support us than to evidence against our views (Ditto & Lopez 1992; Lord & Taylor 2009; Mercier & Sperber 2017). Motivated cognition may even saturate perception. In a classic study, Hastorf & Cantril (1954) had Princeton and Dartmouth students watch film of a football match between the schools. Princeton students perceived twice as many (and more severe) fouls committed by Dartmouth players than did Dartmouth students. More recently, Dan Kahan and colleagues updated this study. They showed their participants video of a political demonstration. One group was told that the demonstration was against the provision of abortion, while the other group was told the demonstration was against the military's "don't ask don't tell" policy on recruitment of gay people, and in favor of a more inclusive recruitment policy. Perceptions of the actions of the demonstrators (for example, whether they obstructed or threatened passers-by) were predicted by the perceivers' prior political outlook, with egalitarians perceiving more aggressive

behavior from anti-abortion protesters than from those protesting the recruitment policy, and vice versa for those with more conservative views (Kahan et al. 2011).

Data like these are often interpreted through the lens of a dual process account of cognition (J. St. B. T. Evans 2008; J. St. B. T. Evans & Stanovich 2013). Type 1 cognition is (always or typically, depending on the version) fast, mandatory in its operation, effortless (both in its phenomenology and insofar as it is not dependent on the availability of cognitive resources), and typically unconscious. Type 2 cognition has the opposite profile: it's slow, must be engaged and sustained effortfully, degrades under cognitive load and is conscious, in the sense that agents know when they're engaged in it. Type 1 cognition is typically assumed to be evolutionarily ancient, whereas Type 2 cognition is a more recent evolutionary adaptation. Type 2 cognition is the kind of cognition we typically associate with intelligence. It's the kind of cognition required for science, math, and philosophy (though Type 1 cognition plays a role in all thought).

Dual process accounts were made famous through the work of Daniel Kahneman and Amos Tversky (Kahneman 2011; Kahneman et al. 1982). Kahneman was awarded the 2002 Nobel Prize in economics; sadly, his co-author died 6 years earlier. Kahneman and Tversky studied *heuristics* and *biases*; rules of thumbs, mental shortcuts and dispositions to weigh information in a variety of ways, all of which make adaptive sense as responses to challenges we faced in the environment in which we evolved, but which may mislead us in contemporary environments. For example, the *salience bias* arises from the way in which information that is easily accessed or emotionally colored is given heavier weight in decision-making than information that is relatively pallid (Bordalo et al. 2010). The salience bias explains why terrorist attacks or mass shootings, which are very salient for us because they're emotionally charged, are apt to be given far more weight than they deserve in our decision making compared to higher probability risks which are much more common and much less salient (Sunstein & Zeckhauser 2011).

Dual process theories provide a neat explanation for our rationality deficits. If the rational response to a problem consists in the response we should reach after careful deliberation, then we should expect departures from rationality to be common. Type 2 cognition is a scarce resource,

one we must use sparingly. Most of our responses will be due to, or heavily influenced by, Type 1 processes, and they will routinely lead us astray. Type 1 cognition may be reliable in the environment in which we evolved, but the contemporary world routinely throws up challenges utterly unlike those it's designed for. We live in groups that are orders of magnitude bigger than those we're adapted for; we're required to make decisions (about retirement savings; about our health; about government policy) that require us to take into account probabilities and effects years or decades into the future, and so on. Type 1 cognition is ill suited to these kinds of challenges, but it remains an important determinant of how we decide. Guided by unreliable cues, we choose badly. We may, for instance, trade away our civil liberties for a tiny increase in protection against terrorist attacks, while ignoring much higher probability (and much more easily addressed) risks like heart attacks and strokes caused by urban pollution.

Bad belief formation might plausibly arise from reliance on Type 1 cognition. It could arise from biased assimilation, for instance. As mentioned above, support for free markets is the best predictor of climate change skepticism. For those who support unfettered markets, climate change is threatening (Bardon 2019; Keller 2015). Because an adequate response seems to require interference with the market, strongly pro-market individuals have a strong motivation to be skeptical. This motivation could help explain their beliefs. They might discount evidence that conflicts with the belief they want to accept, directly or by casting doubt on its source, while being relatively credulous with regard to evidence that supports their views.

Dan Kahan's novel variant on this kind of view has been particularly influential.[10] He explains bad belief formation as a product of *cultural*

[10] Kahan might object to his account being described as a rationality deficit account. On his view, when we ask survey respondents questions like "is climate change real?" we're asking them about their identities, not their views on science: they respond by telling us who they are (this is not expressive responding, as usually understood, because on Kahan's account people report their genuine beliefs, rather than beliefs they don't hold; belief is shaped by identity on this view). But if the task is really protecting, or reporting, our identities, it's far from clear that we act irrationally at all. Nevertheless, it seems appropriate to treat Kahan's account under this heading, both because motivated cognition is one of the mechanisms people have foremost to mind when they postulate rationality deficit accounts of bad belief, and because the mechanism Kahan identifies is supposed to lead individuals to accept false beliefs due to a selective inability to process information as Type 2 cognition would mandate (see Williams, 2021, for further discussion).

cognition. Cultural cognition is motivated reasoning made social: for Kahan, it's our cultural and social identity that shapes how we perceive the world and how we make inferences about it. Kahan sees cultural identities as tending to come in more or less coherent packages. Those who value free markets also tend to be highly individualistic, for example. In the contemporary United States, cultural identities are organized around two principal axes: a hierarchical/egalitarian axis, on the one hand, and an individualistic/communitarian axis, on the other. Hierarchical individualists are the most supportive of unfettered markets, while communitarian egalitarians the least. When our cultural identities are threatened, Kahan suggests, we engage in *identity protective cognition* to defend it (Kahan 2008, 2017; Kahan et al. 2010).

Kahan has produced an impressive array of evidence in support of the cultural cognition hypothesis. One lovely experiment involved participants assessing the efficacy either of a skin cream as a treatment for a rash or of a ban on carrying concealed weapons as a response to crime (Kahan, Peters, Dawson, & Slovic 2017). Participants had to make this assessment on the basis of a 2×2 contingency table. The numbers used were identical across conditions: it was therefore possible to assess the extent to which the topic alone made a difference to participants' capacities to engage in numerical reasoning, independently of the mathematical challenge the tasks involved.

As Kahan and his colleagues predicted, topic made a significant difference to the results. In the two skin treatment conditions, better numeracy correlated with a higher probability of choosing the right answer. That's unsurprising, since the task is moderately difficult and getting the right answer requires comparing ratios to detect covariance between treatment or its absence and improvement. But in the (mathematically identical) gun control conditions, better numeracy didn't predict a higher probability of picking the right response. Just the reverse: more numerate participants exhibited *greater* polarization than those with lower numeracy. This depressing finding—that higher ability may make motivated individuals *less* accurate, not more—is supported by a range of other data. Above, we cited evidence from Kahan himself that greater science literacy and higher levels of education correlate with greater

skepticism about climate change among Republicans (Kahan 2015). This basic finding is supported by other work (Drummond & Fischhoff 2017; McCright et al. 2016). Nurse & Grant (2020) have recently demonstrated the existence of motivated numeracy with regard to climate change specifically and Connor et al. forthcoming replicated the basic finding in a European sample, though they did not find evidence of increased polarization in high numeracy participants.

Kahan suggests that the greater polarization seen among more capable and informed participants is due to their greater capacity. This capacity gives them an ability less capable participants don't possess: to clearly recognize how threatening the correct response is to their worldview, or identity. They are therefore motivated to selectively inhibit Type 2 cognition. Other dual process theorists have suggested different explanations. Perhaps bad beliefs among the cognitively sophisticated arise from the selective *deployment*, rather than inhibition, of Type 2 cognition. Taber & Lodge (2006) suggest something along these lines: having greater reasoning skill and more information available may give motivated reasoners more tools to defend the position they are motivated to accept against unfavorable evidence. Either kind of explanation might explain not only why Republicans who score higher in Ordinary Scientific Intelligence are less likely to accept climate change, but also higher levels of belief that Barack Obama is a secret Muslim among better educated Republicans (Lewandowsky et al. 2012), or why philosophers who specialize in ethics don't behave better than those in other areas of philosophy (Rust and Schwitzgebel 2009; Schwitzgebel 2009a).

Just as information deficits surely explain some cases of bad belief, rationality deficits probably really play a role in explaining some of the kinds of cases we're interested in. But they fall well short of a comprehensive explanation. I'll focus on Kahan's sophisticated version of a dual process account, since he has explicitly developed his theory with an eye toward explaining bad beliefs.

One reason for some skepticism about Kahan's account is the apparent fact that bad belief formation—while systematic in a way that calls for explanation—doesn't always correlate with identity. Skepticism about

climate science and about evolution may be unusual in this regard. In other cases, there is little or no correlation with markers of group belonging. As Kahan himself has noted, anti-vaxx sentiment doesn't seem to be predicted by group identity (Kahan 2014).[11] The same seems to be true for opposition to genetically modified organisms (GMOs) (Lewandowsky, Gignac, & Oberauer 2013). According to Kahan, we engage in motivated reasoning most powerfully when a query is perceived as probing our identities. He seems therefore to be committed to thinking that on GMOs and vaccination sentiment, we should see a close correlation between scientific literacy and accuracy. But that's not the case: anti-vaxx sentiment doesn't correlate with ordinary scientific intelligence.

Further, there are grounds for skepticism about a central plank of accounts that turn on motivated cognition. On these views, we are motivated to reject some hypothesis because it is threatening to our group identity or our self-esteem. Evolution is a clear case in which this might be true: while there are theological views entirely consistent with evolution, theists may be passionately committed to a creation story on which God created the world in seven days less than 10,000 years ago. Evolution is therefore intrinsically threatening to this identity-constitutive commitment. But other cases are much less clear.

An analogous story is often told about climate change: it's inherently threatening to those who support free markets, since any adequate response to it would involve heavy regulation of the market (Bardon 2019; Keller 2015). But it is far from obvious there's any inherent conflict between climate change and the ideological commitments of the vast majority of those who reject it. Ideologies are usually too indistinct to entail or even imply positions on policy. For instance, most people who call themselves fiscal conservatives express as much support for government spending as those who don't think of themselves as fiscally conservative (Barber & Pope 2019; Merkley & Stecula 2018). The gulf

[11] There is, however, some evidence that anti-vaxx sentiment is currently coming to be correlated with right-wing political views (Quintana et al. n.d.). This may reflect the politicization of the COVID-19 pandemic: as a consequence of Trump's opposition to lockdowns, acceptance of public health messaging came to be associated with Democrats and opposition to them with Republicans. Of course, partisan polarization on vaccination beliefs won't save Kahan's account, because he needs to explain why anti-vaxx sentiments were common prior to the polarization.

between our ideological commitments and policy suggests that we rarely reject a proposition due to conflict between them.

Even setting this issue aside, explanations that explain the rejection of a proposition by citing this kind of conflict vastly exaggerate the degree to which there is a fit between particular policies and broad ideological orientations. Adrian Bardon (Bardon 2019), for example, cites the right's commitment to the status quo as a driver. Of course, conservatism is part of the right's brand. But for the past century or more, the right has also been the home of people fervently dedicated to shaking up the status quo. Across the global North, parties on the right are strongly supportive of relatively unfettered capitalism, and capitalism is the most dynamic economic system the world has ever seen and the most corrosive of established practices and institutions. As Marx and Engels (in what might rightly be seen as at least as much a paean to capitalism as a condemnation of it) wrote:

> Constant revolutionising of production, uninterrupted disturbance of all social conditions, everlasting uncertainty and agitation distinguish the bourgeois epoch from all earlier ones. All fixed, fast-frozen relations, with their train of ancient and venerable prejudices and opinions, are swept away, all new-formed ones become antiquated before they can ossify. All that is solid melts into air, all that is holy is profaned.
>
> (Marx & Engels 2012).

In the intervening century and three-quarters, the revolutionizing power of capitalism has only intensified.

The dual commitment of the right to dynamism and to stability ensures that being a Republican has no determinate policy implications. From a contradiction, anything follows. Until fairly recently, in fact, the conservative strain dominated the capitalist strain within the Republican party when it came to the environment. Conservatives might as easily be conservationists as capitalists. Indeed, conservative conservationism has historically been a strong current within conservative thought, and up until recently there was no partisan divide on the environment, either within Congress or among the general public in the United States (McCright et al. 2014). Conservative distrust of environmentalism appears to arise from, rather than cause, the partisan split.

Even now, it's very hard to see right-wing opposition to climate change as genuinely due to ideological commitments. The same right that frets about market interference strongly supports at least $20 billion dollars in annual subsidies to fossil fuel interests within the United States (Environmental and Energy Study Institute 2019). In Australia, in the face of growing competition from renewables the conservative government has floated the idea of subsidizing coal-fired power stations to keep them competitive (Murphy 2020). Absurdly, while the Australian government rejects the idea of a price on carbon emissions due to its negative effects on business, many of the businesses that would allegedly suffer have called for its introduction (Toscano 2018). While the rhetoric might be that climate change is threatening because it interferes with the market, those who engage in this rhetoric are quite willing to interfere with the market to see off the threat.

In a 1988 campaign speech, George H. W. Bush pledged to tackle the greenhouse effect through the "White House effect" (Hudson 2018). While concerns about climate may be very uncongenial to the contemporary Republican party, that fact is probably as much due to the role that opposition to environmentalism played in shaping the ideology of the party over recent decades as to anything intrinsic in protecting the environment. The party might easily have instead followed a path more closely akin to the European center-right, which recognizes the need for constraints on capital in the name of protecting the environment and social institutions.[12] Of course, the left is equally vague in its ideological commitments. The fact that environmental concerns and support for the science of climate change currently find a more congenial environment on the left is a historical accident, not a reflection of the nature of either ideology or the people who hold them.

Defenders of motivated cognition of bad belief formation might shrug their shoulders in response. It's an interesting question, they may

[12] That said, Republican ties to big business certainly played a role in slowing down and even reversing gains on the environment on more than one occasion. Bush rapidly lost enthusiasm for tackling climate change once elected. Oreskes & Conway (2011) tell the story of how people close to the Reagan administration deliberately distorted the science to prevent action on CFCs to tackle ozone depletion. Whether these problems are best seen as arising from partisan ideology or the influence of big business over policy is an open question, however, in light of the failure of Democrat administrations to tackle climate change effectively.

say, *why* people are passionately motivated to defend certain hypotheses, but that's a question for historians, or political scientists, or sociologists. We can help ourselves to the fact that they *are* passionately motivated to defend them, without needing to know why. In later chapters, I'll suggest this response won't do. In fact, people *aren't* passionately committed to defending much of anything. Given the right nudges, they'll passionately defend positions diametrically opposed to those they formerly espoused.

I don't take any of these points to be decisive. For a start, I haven't attempted anything like a comprehensive survey of even the major highlights of the literature explaining bad beliefs within a dual process framework (for instance, I've said nothing about the important work of Gordon Pennycook and David Rand (2019), who explain some cases of bad belief by reference to cognitive laziness and are compelling critics of Kahan's work (Tappin et al. 2020)). Theory choice is a comparative affair: sometimes the best reason to reject a theory is because there's a better one available. Developing such a theory is my task across the remainder of this book. On the account I will develop, people are far more responsive to genuine evidence than the dual process hypothesis concedes. Rather than deploy mechanisms that respond to cues *rather than* evidence, I will argue we respond to cues *as* evidence, and we do so rationally. We are trying to *find out*, not just to fit in.[13]

I won't immediately present the main lines of the alternative account I'll defend. There's more background we need, and this background (too) comes from cognitive science. We are rational creatures and we are social creatures: these two aspects of our being are not independent but intertwined. In responding to social cues, we respond to reasons. The nature of these cues and our responses to them is the topic of the next chapter.

[13] De Cruz (2020) argues for a view she presents as midway between Kahan's and mine. She suggests that while Kahan does not give sufficient weight to epistemic factors, I don't give sufficient weight to the motivation to belong. Some of the evidence she cites does testify to the power of this motivation. For instance, in Asch-style paradigms, people often conform behaviorally without changing their beliefs (see Mercier (2020) for further discussion). However, other considerations she takes to be non-epistemic (e.g., markers of group belonging like accent) I will argue are actually evidential: they turn on higher-order evidence.

2

Culturing Belief

What kind of being are we? This of course is one of the oldest questions in philosophy. In earlier eras, answers were often non-naturalistic (we are animals with souls, for instance). Today, one of the oldest answers is also one of the most popular: with Aristotle, we often think we are distinguished from other animals by our rationality.

Looking for necessary and sufficient conditions is a fool's errand, and a search for essences even more so. There's no property that distinguishes all and only human beings from other animals, beyond facts about descent. But there's something right about the claim that we're rational animals. Our intellectual capacities and achievements are distinctive and impressive. At the same time, there's also something misleading about the common picture of ourselves as rational animals. We think of our rational capacities as realized by our big brains, and there are quite a few grains of truth to that thought. But our rationality also depends on our sociality, and thinking of ourselves as cultural animals is no less accurate than thinking of ourselves as rational animals.

Nineteenth-century history provides excellent, if sometimes grim, illustrations of how limited individual cognition is compared to cultural knowledge. Big brains, good education, and excellent preparation often weren't sufficient to allow nineteenth-century European explorers to survive in difficult environments. In 1846, two ships commanded by Sir John Franklin, on an expedition to chart the Northwest passage, became stuck in sea ice in the Canadian Arctic. The entire crew perished. But the area was regarded by the local Inuit people as rich in resources. Despite their training and experience (Franklin was on his fourth Arctic trip) and ample resources, they were unable to acquire the skills they needed to survive. A few decades later, Roald Amundsen spent two winters in the same region. He relied on the help of the Netsilik Inuit for his survival (Boyd et al. 2011).

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0002

In a very different environment, the Burke and Wills expedition to cross the Australian continent also suffered catastrophe due to rejecting indigenous knowledge (Burcham 2008). Running low on food, members of the expedition accepted a gift of cakes made from the nardoo plant for sustenance. However, apparently as a consequence of unease with being reliant on people they saw as inferior, they spurned further assistance and attempted to make the cakes themselves. They ground the seeds into a powder, mixed it with water, and baked it. Unbeknownst to them, the local people roasted the seeds prior to grinding. This step is required to remove toxins from the plant. Because they missed this step, the explorers didn't get the nutrients they needed from the nardoo cakes. There was only one survivor: he accepted further aid from the Yandruwandha people.

Outback Australia and the Canadian Arctic are harsh environments. But indigenous people didn't just survive in these environments; they flourished. They flourished not only due to their big brains—which they shared with the explorers—but because of their cultural knowledge. Life in the Arctic depends on a rich range of accumulated innovations (Boyd et al. 2011; Richerson & Boyd 2008). Traditional Arctic life requires knowledge how to make special clothing, to manufacture and use special tools for hunting, to construct snow houses, to build fires (without access to wood!) for cooking and melting water, and much more. Each of these skills is complex and hard to learn. Take clothing. The Inuit stayed warm and comfortable by making clothing from caribou skin, which has better insulation properties than seal or polar bear fur. But not just any caribou skin will do: it has to be harvested at the right time of year, and then prepared by repeated stretching, scraping, and moistening. The hides then have to be shaped in ways that maximize heat retention while allowing moisture to escape. A ruff of wolverine fur, especially selected for length, is then added. Footwear is equally specialized, consisting of *five* separate layers: three different layers of stockings, each with a different design, then two different kinds of boots. The know-how needed to make just one of these items of clothing is difficult to acquire, let alone the whole package. Moreover, individual elements of the package (for example, knowing how to prepare one of the layers of stockings) are often little use by themselves: it is only when they play a role in the entire package that they make a significant difference to survival.

If each element of the package is hard to acquire, and many elements have little value on their own, how did the Inuit succeed in acquiring the whole package? Almost certainly, the answer involves multigenerational accumulation of knowledge, with many different group members each playing their small part in the acquisition of the propositional and procedural knowledge of the group as a whole. Cumulative culture, an irreducibly and deeply collective enterprise, is essential to the acquisition of the kinds of knowledge that allows human groups to flourish in highly diverse environments, from the deserts to the tundra and the tropics.

The fate of the Franklin expedition stands testament to how far this kind of knowledge exceeds the ability of a group of individuals to reconstruct from scratch. The Inuit *themselves* found it impossible to reconstruct their own knowledge when it was lost. An epidemic seems to have struck the Polar Intuit sometime in the 1820s, resulting in the death of many older members of the community; that is, they lost many repositories of cultural knowledge (Boyd et al. 2011). As a result, they lost important skills, and the group entered a decades-long decline in numbers. This decline was halted only when they reacquired the skills from another group around 1862. They were not able to reinvent the lost skills of kayak and snow house design in the intervening decades, despite strong motivation and possession of a suite of related skills.

Cultural knowledge solves problems that are intrinsically difficult. When feedback is quick, individual cognition is often up to the task of solving problems. We rapidly learn to avoid suspension-destroying potholes in the road or nettles in the bushes. But when the relationship between an action and its effects is slow to manifest and probabilistic, individuals do very badly on their own. Think of how long it took to demonstrate the effects of tobacco on health; for decades, people denied the link between smoking and cancer, because they were more impressed by salient cases of individuals who had lived to ripe old ages despite smoking heavily (and, of course, because merchants of doubt deliberately muddied the waters; Oreskes & Conway 2011). Science has developed mathematical tools for detecting signal in the noisy relationship between variables, like the relationship between smoking and cancer. Without such tools, individual cognition is highly unreliable. But cultural cognition often succeeds in identifying the signal amid the noise without the need for statistical tools.

Detoxification methods like those developed by the Yandruwandha people (and which the Burke and Wills expedition failed to imitate) are good illustrations both of how difficult these problems can be, and of the spectacular success of cultural cognition at solving them. Because edible plants often evolve toxins as protection against herbivores, many staple foodstuffs are (or were, before very intensive selective breeding) toxic. Take corn. Corn is cheap to produce and high in energy. For this reason, it was exported from the new world, to which it is native, to the old quite rapidly after the arrival of the Spanish. It came to be an important food crop in Italy, Spain, and later the southern United States. But with corn consumption came pellagra. Pellagra manifests first as a skin disease, but untreated it can lead to dementia and even death.

Medical professionals quickly realized that there was a link between pellagra and corn consumption, but suspected some kind of contamination as the cause. It wasn't until the second decade of the twentieth century that the real cause was identified: niacin deficiency. Yet pellagra was very rare among the indigenous people who had relied upon corn for centuries. They avoided pellagra by cooking corn together with an alkali, which releases otherwise chemically-bound niacin. They used wood ash, or ground sea-shells, or lime, depending on what was available locally. But of course they had no concept of niacin or alkali. Asked *why* it is necessary to mix wood ash with corn meal, indigenous people often had no more to say than "it is our custom" (Henrich 2015). They may not have known *that* it is adaptive, let alone *why*. They were and are smart, big-brained primate like the members of the Burke and Will expedition, but it wasn't their individual cognition (alone) that had allowed them to develop the detoxification processes that helped them avoid pellagra. It was cultural evolution, which enables the detection of a signal in a very noisy background.

Cultural Evolution

We started this chapter by asking what kind of being we are. I suggested that an answer as good as any would be that that we're cultural animals. "Culture," as I use the term here, refers to information that is acquired from others, by vertical or horizontal transmission (i.e., from elders or

peers) and which affects behavior. In this sense, culture is not unique to human beings. Famously, when a young Japanese macaque dubbed Imo learned to wash sweet potatoes left for her troop on the beach by dunking the potatoes in the sea, the behavior spread throughout the troop. Since Imo's innovation was first documented, a variety of other behavioral traditions have been observed among macaques (Laland 2017). Chimps have traditions that differ from troop to troop, ranging from different techniques for termite fishing to the use of stone tools to break open nuts; orangutans also use tools, with the tools and the method of exploitation differing from area to area (Schaik et al. 2003).

Human cultures differ from the cultures of other primates in complexity, of course. But more importantly, only human culture appears to be *cumulative* culture. In our species (perhaps alone; certainly to an extent that is dramatically greater than in any other), cultural innovations are not merely transmitted: they become a platform on which others can build. Human culture is subject to the "ratchet effect" (Tennie et al. 2009). The behavioral traditions seen in other primates and other animals can preserve individual innovations and transmit them to future generations, but only cumulative culture builds on these innovations, enabling cognitive achievements that go beyond what any individual or any generation can achieve. It is this development over time that allows cultural evolution to detect signal in noise when the noise exceeds the capacity of unaided individual cognition to parse. It also enables the detection of temporal fluctuations that exceed living memory (Shea 2009). Cumulative culture opens up horizons for knowledge that are closed to individuals, no matter how individually gifted they are.

There's good reason to believe that the mechanisms underlying cumulative culture are evolutionary. Evolution is substrate neutral: it needn't be limited to biological reproduction. Evolution occurs whenever (roughly) there is selection between individuals which vary in their characteristics and this variation is heritable. So long as traits are differentially rewarded, these traits are heritable, and the environment is sufficiently stable over time, we should expect evolution. For instance, prior to (and as a condition of) the emergence of life, evolutionary processes account for the emergence of organic compounds (Bada & Lazcano

2009). Evolutionary mechanisms also account for changes in human behavior across time, independently of changes in gene frequencies.

Cultural evolutionary theory is sometimes mistakenly identified with *memetics*. Memes, first proposed by Richard Dawkins (1989), are units of culture analogous to genes. Memes are subject to selection pressures and they are heritable. Whereas genes (almost always) get selected *en masse*, however (when organisms are selected), memes are selected one by one. They can be fit even when we, their hosts, are not—even when they *lower* our fitness. A classic example of a meme that replicates independently of its fitness effects on its host is an earworm: a catchy tune that gets lodged in our heads long after we hear it. Notoriously, earworms may be disliked by those who experience them, but they are good at replicating themselves. A person who dislikes an earworm can nevertheless be a vector for its replication: she may find herself humming it, for example, thereby contributing to its spread. If she loses friends as a result, we may see a dramatic dissociation between the host's interests and the interests of the meme.

Perhaps memetics explains some features of culture. However, it can't explain the emergence and transmission of cultural practices of any great complexity. The kind of cultural evolution I'm invoking is of a different sort. While it's non-genetic, its effects are primarily on the fitness of the organism (and perhaps the group), not on the fitness of the units of culture—if there are any, in any meaningful sense—themselves. Beliefs, technologies and practices make an obvious difference to our fitness (given that they make a difference to how we behave), and therefore affect our biological fitness. Believing that *that* plant is edible and *this* is not may be the difference between life and death; inheriting the capacity to make a boat of a particular kind may enable a better catch, which in turn enables the person to support more children. Even songs or rituals may make a difference to biological fitness, say by increasing bonds of solidarity that enable a group to avoid open conflict.

Cultural evolution produces adaptive changes in practices or beliefs without (or independent of) changes in gene frequencies.[1] One group

[1] Note that gene-culture coevolution may occur. The classic example is the development of lactose tolerance. A cultural innovation—dairy farming—brought about selective pressure for genetic evolution in those groups to which the practice spread (Cavalli-Sforza et al. 1994).

may be fitter than another without the groups differing genetically; the fitness may instead be due to cultural practices. Cultural evolution has likely been more significant than genetic evolution as a factor in human evolution over the past 50 millennia or so, and perhaps much longer, if only because it can occur very much faster than biological evolution (Perreault 2012). It usually takes many generations for biological evolution to occur, but cultural evolution can occur within a generation.

Some otherwise puzzling facts about human beings might be explained by the centrality of cultural evolution to our flourishing. Human beings have unusually long periods of dependency on caregivers. This long period of dependency is also an apprenticeship in the local culture. We live in a dizzying diversity of environments, and what is adaptive in one may be highly maladaptive in another. The beliefs and behaviors that are adaptive in the Amazonian rainforest are very different from those that are adaptive in the Arctic, which are different again from those required in Karachi or in Copenhagen, or in sub-Saharan Africa, or the Australian outback. Genetic evolution can transmit behaviors in animals (nest building, hunting techniques, song patterns—though even in these cases some kind of learning usually plays a role too). It can even transmit conditional behaviors: that is, it can encode instructions for one set of behaviors in one kind of environment and a quite different set in another (so-called facultative adaptations). But it can't encode for the enormous diversity of complex behaviors needed for flourishing across the range of (ever-shifting) environments in which our species lives. Hence the long period of apprenticeship: we need the time to acquire the set of behaviors we'll need in the specific environment we're born into.

Our early malleability and long dependence isn't the only adaptation we have for the acquisition of culture. We have a whole suite, whether as a product of genetic evolution or (as Cecilia Heyes (2018) has argued) due to cultural evolution itself.[2] A well-known (albeit somewhat controversial) example is our disposition to imitate. We use the verb "to ape" to

[2] Even on Heyes' account, we have some dispositions that are not themselves explained by cultural evolution but which facilitate it. For instance, we are peculiarly tolerant of others and especially the young, allowing them to observe our behaviors.

describe slavish imitation. In fact, no other animal—not even the other apes—apes as much as we do. There's experimental evidence that we're *overimitators*: whereas other animals copy behaviors when they recognize they're instrumentally rational, human beings are disposed to copy even those components of behavior that don't appear to be required for goal pursuit. Nagell, Olguin, & Tomasello (1993) demonstrated a novel technique to human children and chimps. They used a rake, *tine side down*, to draw sweets that were otherwise out of reach toward themselves. Using a rake that way is very inefficient: many sweets slip through the gaps in the tines. Given the opportunity to perform the task themselves, chimps flipped the rake so that the flat side acted as a more efficient tool, with fewer sweets escaping. But human children tended to imitate the action just as demonstrated.

Later experimental work demonstrated that the disposition to overimitate is selective: children overimitate when the behavior appear to be intended by the model, regardless of whether they see the point of the actions. For instance, while children who observed a model turning on a light by butting it with her head tended to do the same, rather than use their hands (Meltzoff 1988), imitation dropped significantly if the model's hands were occupied, suggesting that the decision to use her head was not a component of how things are supposed to be done (Gergely et al. 2002). This sensitivity to whether the behavior is intentional allows children to distinguish between those behaviors that are constitutive of the culture they are acquiring and those that are incidental.

It's worth pausing to appreciate the (prima facie) oddity of this contrast between chimp and human behavior. We humans pride ourselves on our intelligence; we're supposed to be rational animals. Yet when humans and chimps are confronted with the task of procuring a valuable good using a tool, it's chimps and not us who analyze the causal structure of the task to accomplish the goal more efficiently. Children copy the behavior demonstrated if it appears intentional; chimps drop irrelevant features or inefficient techniques. We are *over*imitators: we imitate more than seems justifiable given our goals and the nature of the task. Even when some steps are very obviously irrelevant, and even in the face of prompting to drop superfluous elements, children are reluctant to deviate from the demonstrated sequence of actions (Lyons et al. 2007).

Nor is overimitation confined to children: Flynn & Smith (2012) found that only when adult participants were told that a demonstrator was themselves a novice did they not overimitate.

Why are we prone to ape more than apes? If we're so smart, why do they seem to outperform us in identifying efficiencies and more successful routes to a goal? It's because *we* are cultural animals, and *they* are not (not at least, to anything like the extent we are). Imitation is an adaptation for culture. It allows us to acquire knowledge and practices developed by multiple individuals, individuals dispersed across space and time. It allows us to acquire, and then to build on, deeply social knowledge: adaptive behavior that could not have been developed by any individual de novo, no matter how gifted and insightful that person might be. Sir John Franklin and his party could not develop the techniques that would have allowed them to survive in the Arctic: this suite of techniques and this knowledge must be developed and refined across multiple generations and multiple individuals. What they lacked wasn't intelligence or physical capacity. They lacked the requisite culture: it is only by appropriate imitation that we can acquire the requisite knowledge, and that usually takes long enculturation.

Deeply social knowledge and practices may be partially opaque to those who inherit them. They may be deployed by people who know *that* and *how* they are to be used, but who have mistaken ideas about how they work, or no ideas at all. As we saw, indigenous Americans often gave ethnographers no further justification of the practice of preparing corn with an alkali beyond "it is our custom" (Henrich 2015). Sometimes practices are justified in supernatural terms. For example, Naskapi hunters decide where to hunt by using the shoulder blade of a caribou, heated over hot coals so that it cracks and burns, as a kind of map. There's evidence that this kind of divination is adaptive because it effectively randomizes behavior, overcoming our disposition to detect illusory patterns (Henrich 2015). This kind of adaptive use of divination is widespread across multiple cultures. In yet other cases, behavior is given a naturalistic but false rationale. For example, the more toxic marine foods are tabooed for pregnant women in Fiji, significantly lowering their risk of miscarriage (Henrich & Henrich 2010). But justifications for the practice offered range from custom to the idea that the

child would take on the properties of the animal eaten (e.g., that it might have rough skin if its mother ate shark).

Causal analysis, chimpanzee-style, is not required for the acquisition and deployment of cultural knowledge. In fact, since the products of cumulative culture are complex and the contribution of some steps obscure (why roast the nardoo seeds? why add wood ash to the corn meal?), such analysis risks degrading the value of cultural knowledge. The cultural apprentice does better taking the technique on trust. *That's* how we do things, so that's how I'll do them. Second guessing the technique is appropriate when it's a component of shallow cultural knowledge, like the behavioral traditions seen in other primates: when what is transmitted is an innovation that is within the cognitive grasp of a single individual, tinkering with it may easily reap rewards. When what is transmitted is deep cultural knowledge—when the culture is cumulative culture, and successive innovations have become the platform for further elaboration—then it is maladaptive to attempt to innovate by analysis. *Deference* to custom is the appropriate attitude.[3] We might say that we owe our success to the fact that we are in some ways less—or at any rate less directly—rational animals than chimps. We defer to tradition (relatively) unthinkingly, in conditions in which they would analyze causal structure and innovate (in later chapters, I'll suggest that this deference is more rational than it might seem).

Our disposition to overimitate is just one of our adaptations for the acquisition of local norms, practices, and conventions. Conventional ways of behaving are essential to the coordination of action in complex

[3] Of course, too much and too effective deference would put an end to cumulative culture by preventing further innovations. To be cumulative, innovation must occur. It remains somewhat mysterious how we pull off the difficult balancing act required—faithful transmission combined with some degree of innovation—insofar as the two elements are in conflict with one another (Sterelny 2012). Fridland (2018) argues that imitation supports faithful transmission and explicit pedagogy supports innovation. As she notes, when skills are causally opaque, imitation is required for transmission while innovation threatens to be fatal. But when skills have a causal structure that can be discerned, we may do better to transmit them by breaking them down into steps. This proposal leaves the central problem unresolved, however: how do causally opaque complex practices develop in the first place? Fridland's proposal seems to require that such practices must first develop through explicit pedagogy, and subsequently become the target of imitation. But that proposal cannot account for deeply social practices, which have components that those who transmit them may not be able to grasp. *Pace* Fridland, some kind of more genuinely evolutionary process, involving mutation and selection, must play a significant role in the production of deeply social knowledge.

societies, but have an arbitrary content (D. Lewis 1969). Think of driving conventions: it doesn't matter whether we drive on the left or the right, but it matters very much that we all agree on which side we drive on. Because conventions are arbitrary, we can't guess or infer what the local convention may be. It's not only conventions that differ from place to place; so do norms, some with very complex contents (think of how whole books get written about norms of etiquette, not to mention moral norms). The *conformist bias* is a disposition to acquire the local ways of behaving, enabling us to acquire the right set of conventions and norms (Henrich & Boyd 1998). The conformist bias enables us to pick up adaptive ways of behaving without paying the costs of exploration.

While the conformist bias helps us to acquire the local conventions and norms, the *prestige bias* leads us to imitate particularly successful individuals (Chudek et al. 2012; Henrich & Gil-White 2001). The link between their success and their behaviors may be opaque to us. Greater success could be due to their hunting techniques, their social network, their religion, their diet, their dress, and so on. Since the relationship between success and behavior is often causally opaque, we do better to copy successful individuals relatively undiscriminatingly. The prestige bias (like all heuristics and biases) leaves us vulnerable to certain sorts of exploitation: advertisements that link a celebrity or a well-known athlete to a watch or a fragrance take advantage of our disposition to copy successful individuals, even when there's no apparent link between their success and the behavior imitated. But it also helps to drive the evolution of behavior by allowing more successful behaviors to spread within and across groups. I'll call the prestige bias and the conformist bias instances of *social referencing*—looking to others within our social group for cues for what to believe and how to behave. As we'll see in later chapters, social referencing is a very widespread means for people to acquire and update their beliefs.

Imitation, the prestige bias and the conformist bias are adaptations for culture that help to explain why we're less discriminating in what we copy and less disposed to analyze imitated behaviors than chimps. Just how discriminating the mechanisms underlying cultural evolution are is controversial. I've presented a picture that leans very heavily on one

side of these debates. The scientists I've cited (Peter Richerson, Joe Henrich, Robert Boyd) are sometimes said to constitute the Californian school (Sterelny 2017). The so-called Paris school (led by Dan Sperber) takes a different view on many issues, and in particular on how discriminating the mechanisms of transmission are. This isn't the place to address this dispute in any detail, but I will say a few words in justification of my reliance on the Californians (but—importantly—not in opposition to the Parisians; the two are far more compatible than the polemics suggest).

Sperber and his colleagues argue that relatively undiscriminating imitative mechanisms play a smaller role in the transmission of culture than the Californians suggest. Instead, the Parisians argue that culture is reproduced largely through mechanisms that are *reconstructive*, with the learner contributing a great deal to the final product. The Parisians maintain that the reconstructive mechanisms are intelligent and strategic, rather than unthinking and deferential. For example, they cite evidence that children and adults are far from undiscriminating in whom they imitate (Mercier 2017). We are also highly selective in whose testimony we accept. We filter out testimony from out-group members, when there is evidence that it conflicts with the consensus, and when it comes from those that have track records of unreliability or who have shown lack of benevolence to us in the past (P. Harris 2012; Mascaro & Sperber 2009; Sperber et al. 2010). In doing so, we filter testimony by reference to cues that correlate with reliability: consensual testimony is (other things being equal) more likely to be accurate than dissent; malevolent testifiers are more likely to mislead us than benevolent, and so on. Filtering testimony in these ways is the behavior of a rational animal, not an unthinking conformist.

In emphasizing the intelligence of the mechanisms underlying the transmission of culture, the Parisians offer an important corrective to the perception that it is undiscriminating and unthinking. But it's the *perception* they correct; intelligence doesn't enter the picture only when Paris-style mechanisms are in play. Imitation, California-style, is not reflexive and automatic. Instead, it manifests a great deal of intelligence (see Boudry 2018; Buskell 2016). In fact, even our apparently automatic

imitation *itself* manifests intelligence (it's a major aim of this book to show that's true). We don't face a choice between Parisian intelligence and Californian automaticity: the mechanisms emphasized by both sides should be seen as intelligent.

A second reason to think that the conflict between Paris and California is smaller than is sometimes thought is that the two schools focus on the explanation of different aspects of culture. The Parisians' focus is on cultural lineages (see, especially, Morin 2016). High-fidelity imitation is not necessary for the preservation of these lineages: folk tales, painting styles, and myths. In these domains, reconstruction (coupled with what the Parisians call "cultural attractors") may suffice for the transmission of culture.[4] But unobvious technological innovations and sophisticated environmental knowledge—the kind of knowledge that depends on the detection of a signal in a noisy environment—do depend importantly on such imitation.

Agents, adults as well as children, default to imitation when success depends (or seems to depend) on following a precise sequence of steps, when mechanisms are causally opaque, and when the demonstrator is presented as an expert (Acerbi & Mesoudi 2015). It is in these sorts of contexts that reconstruction risks the loss of cultural knowledge. If agents intelligently reconstruct the causal process, they may easily leave

[4] Cultural attractors are species-typical cognitive dispositions that play an important role in stabilizing cultural traditions. Because certain features of a narrative (for example) come naturally to creatures like us, transmission from teacher to learner can be noisy: the novice will reconstruct ambiguous or partial information in ways that match the transmitted narrative except in the (unlikely) event that the narrative is excessively unintuitive. But in the domain of deeply social knowledge, reliance on such attractors would be risky and therefore is likely to be minimal. Such attractors risk distorting the knowledge transmitted, because deeply social knowledge is often highly *un*intuitive. It is also worth noting that the cultural attractor story itself plausibly requires culturally specific, rather than species-typical, machinery to explain content. Take minimal counterintuitiveness theory (e.g., Atran 2002; Boyer 2000), often hailed as one of the great success stories of the approach. According to this account, minimally counterintuitive entities have an advantage when it comes to cultural epidemiology: they are more likely to be recalled and more likely to be transmitted. However, Sterelny (2018) has persuasively argued that the institutional and ritual scaffolding of religious belief is central to its survival. These factors are not species-typical, of course—not in their contents, at any rate—and they are neither invented by individuals nor understood by them as belief transmission mechanisms. If we can generalize the lesson, the attractors themselves may turn out to be far more deeply cultural than the Paris school would be comfortable with.

out crucial steps, since their contribution is often opaque to us (recall the fate of the Burke and Wills expedition). While the telling of folk tales may be adaptive, their precise content typically doesn't matter much, and reconstruction can be given free reign. But when the precise content is crucial, deference is required. We're therefore sensitive to cues for switching from reconstruction to imitation. These cues include the presence of experts, causal opaqueness, ostensive communication ("look at what I'm doing here") and other cues that indicate conventionality (Acerbi & Mesoudi 2015). In a slogan: *deeply* social knowledge depends on deference. In some domains, reconstruction—Paris-style—is probably the primary means of cultural transmission. But in the domains of causally opaque technology, such as food preparation techniques, and the detection of signals in noisy environments, imitation reigns supreme.

Up to this point, we've focused on cultural knowledge: knowledge of the behaviors that we need to function as members of a particular society and to flourish in sometimes harsh environments. The mechanisms we've examined have the function of enabling us to distinguish signal from noise in causally opaque systems, or to identify regularities at temporal and geographical scales that exceed the grasp of an individual. I noted that prior to the development of statistical tools, these mechanisms were the only means we possessed for the detection of such signals. Now, of course, we possess the tools of science, which allow us to achieve the same sorts of ends much more quickly, efficiently, and accurately. These tools can be deployed by individuals. Does that entail the end of the millennia-long age of deeply social knowledge?[5] The next section will assess the extent to which knowledge is social today, in environments far removed from the environment of evolutionary adaptiveness.

[5] Roughly speaking, the claim that knowledge of the causal regularities essential to our flourishing has transitioned from being deeply social to being individual inverts the major claim of Elijah Millgram's *The Great Endarkenment* (Millgram 2015). Millgram argues that as more and more domains of knowledge becomes hyperspecialized, our capacity to understand specialist knowledge shrinks and knowledge becomes fragmented. Millgram provides good reasons to reject the idea that knowledge has recently become less social. But he dramatically overestimates how individual it was prior to the age of hyperspecialization. Human knowledge has *always* been deeply social, and individuals have never had the capacity to genuinely grasp the kinds of knowledge that have always been distinctive of our species.

Science on Mars

Until the last decade of the twentieth-century, epistemology was a largely individualistic enterprise. It was primarily recognition of our pervasive reliance on testimony that altered the landscape (C. A. J. Coady 1992; Lackey & Sosa 2006). We are dependent on testimony, implicit and explicit, for our knowledge about the temporally and geographically distant, and for much of our knowledge about the unobservable posits of science. We learn the history of our country and about major events in the past few centuries via explicit testimony in schools and colleges, and about political events, current affairs, celebrity gossip, and much more through media (increasingly, social media). Sources of implicit testimony include fictional narratives (which encode a lot of information they don't explicitly assert: for instance, about social norms, or about kinds of things and states of affairs) as well as all the subtle cues which convey to us how we should speak, behave, and even think.[6] All this knowledge is social: it's conveyed to us by others and largely taken on trust. For much of what we know about the world, we are deeply dependent on others.

Science, however, may seem crucially different. Doesn't science demand we take nothing on trust? The motto of the Royal Society, the oldest national scientific society in the world, is often taken to express

[6] Epistemologists may quibble with my very expansive use of "testimony" to describe how information is conveyed from one or many agents to others. Certainly implicit testimony often fails to satisfy *any* of the criteria put forward by Coady (1992: 42): for instance, it may not be directed toward those who are in need of evidence. A case can be made that implicit testimony doesn't even satisfy the criteria set down by Lackey's (2008: 35–6) more permissive disjunctive account of testimony, though this is a harder issue to judge (much depends on how we interpret "is reasonably taken as conveying that p"). I don't intend to offer a definition of testimony. It's enough for me that the instances of what I'm calling implicit testimony here clearly have features that warrant treating them as or alongside testimony: they involve the transmission of information from one set of agents to another, whether the first intends to convey the information or not, and independent of whether the second set of agents is conscious that information has been conveyed to them. The information thus conveyed functions as evidence for the second set of agents and they respond to it accordingly. To me, these features seem sufficient to justify calling these instances of information transfer "implicit testimony." If the reader prefers some other term, so be it.

the scientific attitude: *Nullius in verba*—take no one's word for it. Science seems to require that every claim be regarded with a skeptical eye, and nothing be accepted unless and until it has been adequately tested. In science, our naked intelligence and our capacity to test the facts is all that really counts—right? Call this the *Martian model* of science, after the film (and novel) in which a lone individual's ability to "science the shit" out of things enables him to survive in an unforgiving environment, and call intuitions that accord with this model *Martian intuitions*.[7]

Martian intuitions are powerful, and not only as an explanation of the success of science. But they are misleading. Under a variety of (reasonably undemanding) conditions, group deliberation outperforms individual deliberation (Mercier & Sperber 2017). Consider our performance at reasoning tasks like the Wason selection task. In this task, participants are presented with four cards, and asked which cards must be turned over to test if some rule is being violated. The rule is a conditional: if *p* then *q*, and the cards represent *p* or not-*p* on one side and *q* and not-*q* on the obverse. For example, the rule might be "if there is a D on one side of a card, then there is a 5 on the other side," and the cards presented might be the following:

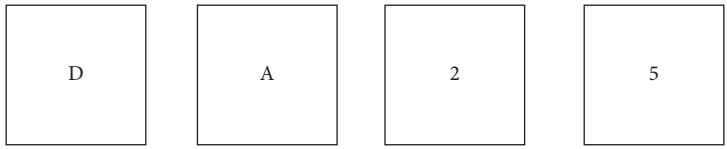


Figure 2.1 The Wason Selection Task.

[7] The line "I'm going to have to science the shit out of this" occurs in the film version of *The Martian*, but not the original book. It is, however, an accurate encapsulation of the flavor of the book as well as the film. Interestingly, the film and the line resonated strongly with many scientists (Neil deGrasse Tyson tweeted that that was his favorite line in the movie), despite the fact there's good reason to think that the success of science is explained by its structure and the division of cognitive labor it exhibits—by the way in which it instantiates features of the cultural model, the rival of the Martian model—rather than by the way it instantiates the Martian model.

Though the task is logically simple, most people do badly at it: only about 10 percent of people select the right cards on the task (in this instance, the first and third).[8] But *groups* of individuals do very much better. In fact, groups of deliberators may manifest the *assembly bonus effect*, where the group performs better than the best individual within it.

Despite the better performance of groups on the Wason selection task, and on many other reasoning problems (see Sunstein 2006; Surowiecki 2004 for accessible reviews), the Martian intuition remains powerful. Groups have a bad name (think of books with titles like *Extraordinary Popular Delusions and the Madness of Crowds* (Mackay 2011)). We badly underestimate group performance on reasoning tasks; we often think groups will do worse on such tasks, rather than better. Academic psychologists with a specialism in human reasoning, who are well aware of the dismal individual performance on the Wason selection task, underestimate the benefits of group deliberation on the task to the same extent as do laypeople. Managers of teams, individuals from East Asia and WEIRD people, all alike underestimate the benefits of group deliberation (Mercier, Trouche, Yama, Heintz, & Girotto 2015).

We are, it seems epistemic individualists who are reliant on social networks and culture for our epistemic success. What explains our individualism, in the face of the pervasiveness of epistemic dependence? Our obstinate individualism may be partially (and only partially) due to the fact that it is sometimes epistemically fruitful to insist on one's private evidence. In fact, some degree of epistemic individualism might be conductive to *group* deliberation. Group deliberation is subject to characteristic limitations and pathologies. Information cascades can mislead the group; powerful individuals can carry disproportionate weight and people may self-silence in the face of prejudice or anxiety. All of these

[8] One common mistake consists in picking cards representing the consequent (in this example, the 5). That's a mistake, because the rule is a conditional and not a biconditional. It might therefore be thought that people are guilty not of a reasoning error but of misunderstanding the nature of the task. One problem with this suggestion is that individual performance on the Wason selection task improves dramatically if the rule is a social contract rule (Cosmides & Tooby 1992). Since changing content while keeping the instructions fixed improves performance so significantly, it seems that the original results are not driven by a confusion brought about by the wording.

problems can be mitigated if people are disposed to give private information and their individual deliberation disproportionate weight.

Take information cascades. Such cascades may occur when agents are led to disregard private evidence because the public evidence outweighs their own. Consider a case like this (based on Anderson and Holt 1997). You're given the opportunity to draw a ball from one of two urns. One urn contains white balls in a ratio of 3:1 to red; the other the reverse. Your task is to identify which urn is which, by drawing a ball from one. Obviously, if you draw first you should bet that the urn you draw from contains predominantly balls of the color you've drawn. Suppose, however, you draw after other agents. You don't know what color ball they've drawn, but you do know how they've betted. With all agents acting rationally, unlucky initial draws can give rise to an information cascade and rising confidence in a falsehood. Suppose the first and the second agent both draw a red ball, and bet accordingly. Their behavior is rational, but their evidence may be misleading: by chance, they've drawn red balls from the predominantly white urn. The agent choosing third in sequence may now rationally bet that the urn is predominantly red, *no matter what color she draws*, because the evidence stemming from the betting behavior of the earlier agents suggests the urn is predominantly red. From this point, the total evidence available to each agent—their private evidence plus the evidence of the betting behavior of the previous drawers—favors red. As the sequence continues, successive agents become increasingly confident that the urn, which is actually predominantly white, is predominantly red; this despite the fact that the majority of their collective evidence supports the correct conclusion. The problem of information cascades can be reduced if individuals are overconfident: if they take their private information to have a disproportionate weight, relative to public information. In some conditions, groups can deliberate better due to the individualism of some of their members.

Even preferring one's own opinions to those of epistemic *superiors* can be epistemically productive, because self-silencing by those who recognize their inferiority may lead to "hidden profiles": information relevant to deliberation going unshared. In fact, groups may benefit from the arrogance of intellectual inferiors even when the information they insist on turns out to be misleading (Surowiecki 2004). Mercier

and Sperber (2017) suggest that some of our individual-level reasoning pathologies may be adaptations for collective deliberation: the confirmation bias, for example, which leads us to overvalue evidence in favor of hypotheses we are well disposed to and undervalue contrary evidence, may conduce to an effective division of epistemic labor, given a diversity of views within the group. Similarly, I suggest, some degree of epistemic individualism—a disposition to underweight the views of others and overweight our own—may be an adaptation for collective deliberation (Levy 2019a).

Whatever the explanation for our epistemic individualism may be, that it's strong—perhaps especially among us WEIRD people (Henrich et al. 2010)—is beyond doubt. We're especially prone to embrace it with regard to science: our cultural paradigm of the scientist is the lone genius, working in isolation in his [*sic*!] lab, often misunderstood or ignored by his contemporaries. But this paradigm is deeply misleading. Scientists, too, are dependent on testimony, explicit and implicit, even *within the domain of their own expertise* (Rabb et al. 2019). Working scientists use tools they didn't develop (and that they may not be able fully to understand), often applied to data they didn't gather and which they can't verify, to test hypotheses that are constrained by theories they may not grasp. These constraints *enable* them to do science. Since climate change denial is a principal exemplar of bad belief in this book, let's take climate science as an example.

Climate science is heavily interdisciplinary: the models and the data that climate scientists use are the product of multiple researchers and research groups across multiple fields. The tools, the techniques, and the fields of expertise of climate scientists vary widely, from taking ice core samples to constructing mathematical models, from studying the physics of air circulation to estimating the extent of previous sea rises from geological evidence. Climate scientists calibrate their findings and refine their models in the light of evidence from other fields. Like every other scientist, moreover, they rely on tools they didn't design and which they lack the skills to fully understand. Central examples include the computers they use, and the mathematical tools they apply to build their models. They rely on these tools, despite having a limited understanding of how they work (Winsberg 2018). Moreover, each climate scientist

works on a narrow area. A paleoclimatologist, for instance, has little expertise in Arctic sea ice or glaciology. But she may depend on experts in those areas for interpretation of some of her data. When it comes to assessing evidence that lies squarely within her competence, she will assess rival hypotheses in the light of prior probabilities that are sensitive to findings of other scientists in other fields (think of how a psychologist might assess evidence in the light of prior probabilities sensitive to information about evolution: if it is difficult to see how a particular psychological mechanism could have evolved, she'll think it unlikely that a mechanism of that kind can explain her observations).

Of course, the spectacular degree of interdisciplinarity seen in climate science is not evident across all sciences. But no science uses only tools it developed, or that its practitioners fully understand. Every science uses mathematical tools. While some capacity to understand these tools is a prerequisite for their competent use, it is routine for many practitioners not to fully grasp their details and near universal for their development from scratch to be beyond their capacity. Working scientists inherit these tools, which often owe their origins to work outside their field. Scientists also use physical tools they could not develop and often do not fully understand. Few neuroscientists can build or even troubleshoot fMRI machines. Less obviously, they rely on physicists and mathematicians for the development of the algorithms used to transform blood oxygenation signals into images of the brain. Every scientist working in the field of biology presupposes the truth of evolution ("nothing in biology makes sense except in the light of evolution," as Dobzhansky (1973) famously wrote). Epistemic dependence is routine in science: dependence on others for data, for tools and techniques, and for theories.

Science produces knowledge not *despite* this dependence, but *because* of it. The spectacular success of science is largely due to its structure as a set of epistemic institutions. This structure plays a central role in neutralizing, sometimes even harnessing, the influence of individual scientists' biases, whether their ideological preferences or the biases identified by psychologists (the confirmation bias, recency effects, the salience bias, and so on). Peer review, for instance, helps to ensure that scientific findings are subject to scrutiny from a variety of perspectives, each

with its own biases. Since these biases often conflict (especially insofar as reviewers may be hostile toward the findings of the paper), their influence is mitigated. The confirmation bias—roughly, our disposition to seek confirming evidence for a hypothesis, and overlook disconfirming—isn't just mitigated: it's harnessed. Because each of us is motivated to search for evidence that favors our preferred theory, together we ensure that all our evidence is brought to bear (Mercier & Sperber 2011, 2017).

Of course, peer review subjects scientific findings to scrutiny by a small number of people, with a restricted range of disciplinary expertise. But the de facto institution of *post* publication peer review is very much more powerful. Every high profile paper is read by experts across a range of fields and is assessed for plausibility in the light of their disciplinary expertise; this collective scrutiny increases the likelihood that flaws in the gathering and interpretation of data are identified alternative hypotheses considered, and overlooked evidence considered. If the findings survive this collective scrutiny, they're transmitted to others and given some degree of credence.[9] Other researchers and research groups may attempt to build on them, or constrain their formulation of new hypotheses and their interpretation of data in the light of the findings. Others, of course, may not be convinced, and instead seek to develop and test alternative hypotheses. Even for these researchers, the findings constitute testimony: they take on trust that the reported experiments were conducted, with the populations described in the methods section, and the results are accurately reported.[10]

Peer review and post publication review is (somewhat) conflictual. But the success of science is also heavily dependent on cooperation. In climate science, we see different disciplinary specialisms tackling a

[9] This might best be seen as a special application of Goldberg's (2010) principle "if that were true, I would have heard it by now"—if there were serious problems with this paper, they would have been identified. Of course, many papers don't receive adequate post-publication review, so the principle applies only to higher-profile papers.

[10] This trust is not undiscriminating, of course. Statisticians have developed tools for the detection of p-hacking; that is, the manipulation of data to produce statistical significance (see, for instance, Head, Holman, Lanfear, Kahn, & Jennions 2015). Since p-hacking involves departures from the advertised methods (for example, by adding additional participants to tip a p-value below significance, or the failure to report unsuccessful trials), these tools ensure that data does not have to be taken entirely on trust.

common problem. Collaboration is also typical at the level of the individual lab or paper. In the sciences, single-authored papers are unusual, and the number of authors per paper is rising (Mallapaty 2018). Co-authors bring different skills and ranges of expertise to papers; often the skills and expertise of different fields. Some authors are engaged in data collection, others in experimental design, yet others in interpretation of the evidence. Some may be experts in techniques that the paper utilizes, and have little or no expertise in the topic itself. They must *trust* one another in order to collaborate, because they always lack the time, and usually lack the expertise, to assess the contribution others make. This is true even if they all belong to the same field. Hardwig (1985) gives the example of a paper in particle physics with 99 authors; each with slightly different expertise. Since then, the trend in physics has been toward much larger groups of authors; the (relatively) recent paper reporting the detection of the Higgs boson had more than 5000 authors (Castelvecchi 2015).[11]

The extent of relatively undiscriminating deference exhibited by scientists is best illustrated by their attitudes to a well-established protocol. They often have little idea of the causal contribution (if any) of every step. Why *that* amount of a solvent, rather than 5ml more, or 5ml less? Why *that* amount of time in the centrifuge? Especially if the experiment is expensive to run (in terms of resources or time), scientists have little incentive to attempt the kind of trial and error experimentation required

[11] Winsberg et al. (2014) worry that deeply distributed science threatens *accountability* for its epistemic achievements. When science is distributed across many individuals and across time, and no one fully understands the methodological choices made, no one "is in an epistemic and social position to explain and defend the methodological and epistemic standards employed in arriving at such a claim, as well as how they were met" (16). They do not, however, give a reason why any individual *should* be accountable in this kind of way. What is lost when accountability goes? This is a pressing question, because (contrary to what they seem to think) the deep distribution of knowledge production is not a recent phenomenon: if accountability requires being able to give an account of how knowledge was generated, or even what is consists in, then it was lost some time in the Paleolithic. Of course there are significant challenges in deeply distributed science, arising from the fact that (as they put it) there is no guarantee that the methodological choices made by the different individuals and groups who contribute to a larger research project, each of which involves epistemic trade-offs, "will form one coherent justification" (19). But that's a question about what they call the *reliability* of the research, not its accountability. It is a pressing question at the level of the individual research project, but in a well-organized and properly stress-tested science (one in which problems are tackled from multiple angles by groups with different interests, aims and agendas), we can expect these issues to be identified and resolved.

to discover whether certain steps might be dropped or compressed. Unless they have a particular interest in some step, they are likely simply to stick to the protocol. As Shea (2009) puts it, "all sorts of techniques and steps are copied without any appreciation of whether or why they are necessary to achieve the goal – following an experimental protocol can feel rather like following a magic spell" (2434). Faced with an established protocol, scientists behave like children shown a technique, rather than like chimps. Rather than innovate, they imitate.

Contemporary science doesn't individualize knowledge; if anything, it distributes it ever more deeply (Millgram 2015). It is because multiple individuals work on a common problem that we make progress on its solution, and these individuals are always themselves embedded in broader research cultures and research groups. The myth of the solitary genius is just that: a myth. In *The Enigma of Reason* (2017) Hugo Mercier and Dan Sperber probe a paradigm of that myth: Werner Heisenberg's withdrawal to a small island to think through deep problems in quantum theory. Heisenberg was, of course, a brilliant thinker. But he did not develop the uncertainty principle on his own. The advances he made were the product of collaboration and dialogue. They emerged through confrontation with the work of multiple other thinkers—Schrödinger, Bohr, and Dirac most prominently, but also other less famous physicists—and through a voluminous correspondence with Wolfgang Pauli. Heisenberg on his island was far more solitary than most scientists ever are, but even he was not alone: not intellectually. It was through dialogue and collaboration that he made his breakthrough.

We are the type of rational animal we are because we are cultural animals. We possess a suite of dispositions that orient us toward others epistemically. For us, knowledge production is in essential respects a product of the distribution of cognitive labor. Distributed cognition is more than merely a very efficient way of exploring the logical and empirical space, and mitigating, or even harnessing, our biases. It opens up new cognitive horizons that would otherwise be entirely inaccessible to us. Science does not free us, the animals we are, from epistemic dependence; if anything it increases it.

3

How Our Minds Are Made Up

In the previous chapter, I surveyed a range of evidence that—I claimed—shows that knowledge production is a highly distributed enterprise, and that each of us is epistemically dependent on others for much of what we know. In the past 30 years, philosophers have come to recognize one facet of this reliance—our dependence on testimony—but have overlooked others. We know a great deal of what we know not only due to testimony, but also due to the ways in which epistemic labor is distributed and due to the ways in which knowledge is implicitly embedded in our social, and even physical, environment. It will be this embedding, and how it drives belief update, that will be the focus of this chapter.

Enlightenment Myths

Here's a caricatured and brutally compressed summary of the received view regarding the past few centuries of Western (*sic*!) thought: In the Middle Ages, intellectuals labored under the shadow of the Church. Only a few brave souls dared defy its orthodoxies. Even Galileo had to recant his heliocentrism in the face of the threat of torture, or worse, while Copernicus' *Des revolutionibus* was banned. Slowly, however, we emerged from this intellectual repression and into the radiance of the Enlightenment. We learned to think for ourselves, rejecting all orthodoxies, all authorities, except the authority of our own reason. "*Sapere Aude*! 'Have courage to use your own understanding!'—that is the motto of enlightenment," as Kant (1784/1991) famously wrote.

This caricature isn't merely simplistic, it's dead wrong. Reliance on our own epistemic powers is epistemically paralyzing. Alone we understand nothing. Knowledge is a social product. In this chapter, I'll focus on our epistemic deference in belief acquisition and update. We defer to

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0003

others so ubiquitously and so routinely we fail to notice when it occurs. We fail to notice our beliefs are dependent on what others believe, and shift as theirs do. We fail to notice this even when the changes are abrupt. These shifts may appear arbitrary and irrational. I'll show that they're neither. In deferring, we follow the evidence, though that fact may not be apparent to ourselves or to others. I'll begin in an apparently unlikely place: support for Donald Trump.

"Never Trumpers" were prominent Republicans who professed obdurate opposition to the candidacy of Donald Trump. They declared him unfit for office, on the grounds of his character and his policies. But most (though by no means all) Never Trumpers withdrew their opposition, and many became fervent supporters. This looks like belief update in the absence of evidence: Trump himself never provided any reason for them to think that their opposition was ill-founded.

Lindsay Graham, a long-serving Republican senator and himself a former candidate for the nomination, is an irresistible example. In February 2016, Graham described Trump as a "kook" who is "unfit for office." By November 2017, he sang a rather different song, now expressing his deep concern at the way the media stooped to portraying the president "as some kind of kook not fit to be president" (Lopez 2017). By April of the following year, Graham had decided that Trump deserved the Nobel Peace Prize for his reconciliation with North Korea.

Of course, Graham's about-face may reflect nothing more than the banal hypocrisy of politics. No one will be shocked to learn that politicians abandon professed convictions for political advantage. In any case, we shouldn't build theoretical castles on individual cases. But this kind of case isn't uncommon. Consider the swift abandonment of strongly held principles by white evangelicals. In 2011, white evangelicals were the group most strongly committed to the view that personal morality is an essential prerequisite for elected office. By 2016, they were *least* committed to that proposition (Kurtzleben 2016). This is surely a consequence of their preference for Donald Trump, a man who was brazenly immoral in his private life. A number of Evangelical leaders who had signed a letter condemning Bill Clinton on the grounds of his personal immorality, arguing that such immorality was more important than policies, now supported Trump and offered his policies as their reason (D. Miller 2018).

Recent examples of dramatic and apparently unprincipled changes in belief tend to occur on the political right. That's probably due to recent trends in world events, not anything intrinsic to being on the right. The candidacy and subsequent election of Donald Trump represented a watershed moment in US (and world) politics, for multiple reasons, but one in particular made such shifts more likely. Like all major political parties, the Republicans represent a broad swathe of somewhat conflicting interests: social conservativism (for instance) sits uneasily with free market fundamentalism, because the market is corrosive of traditional institutions (as we noted in an earlier chapter). Because parties yoke together different interests, candidates typically attempt to represent themselves as sensitive to the aims and desires of all of them. Trump decisively broke with this kind of within-party compromise, rejecting both the commitment to global trade and to at least a veneer of civility and respect for tradition that had characterized the self-conception of much of the party. As a consequence, those who identified as Republicans were forced to choose between party and these values. No doubt if the left was faced with as stark a choice, some would repudiate proclaimed core principles just as quickly.

Sudden shifts like these seem to occur irrationally, and in the absence of any evidence that could justify them. Trump's policies and behavior didn't change over the period during which Graham shifted from determined opposition to fervent support; not in any way that could explain the reversal. I'll argue that shifts like this typically occur as a result of our adaptive disposition to outsource belief to other agents. They manifest our nature as social epistemic agents. This outsourcing, moreover, is *rational* and evidence-responsive: it manifests our nature as *rational* social animals.

Outsourced Belief

In the last chapter, we surveyed the extensive evidence that human flourishing is very significantly due to our capacity to engage in distributed cognition. Human beings are distinctively dependent on cultural evolution for the development of the tools, techniques and practices that

enable us to colonize a dizzying diversity of environments. Cultural evolution generates cultural knowledge. It generates knowledge-how and it generates a great deal of propositional knowledge ("caribou skin must be harvested in the fall, because it's too thin earlier in the year"). Cultural evolution relies on (or equips us with) a suite of adaptations that enable us to acquire beliefs from others, and thereby acquire the knowledge needed for flourishing. We also saw that science builds on these dispositions. It institutionalizes collective deliberation (for example, in the form of peer review), with distributed groups of scientists both deferring to and testing the hypotheses and findings of other groups. It depends on pervasive, but selective, epistemic deference. The Martian model of science introduced in the last chapter is false, at least here on Earth and perhaps anywhere there are beings capable of engaging in science.

Scientists outsource belief production and they outsource beliefs themselves. It is obvious how they outsource belief production: they depend on other scientists, both those working in their discipline and those in other disciplines, for the data and theories that are inputs into and constraints on their work. The outsourcing of beliefs comes in two varieties. First, scientists may rely on others to believe things on their behalf. Second, they may outsource beliefs they themselves hold by relying on others, or even the social or physical environment, to ensure that cues that trigger those beliefs are made available at appropriate times. Scientists outsource beliefs in this kind of way when they work in interdisciplinary groups. Science institutionalizes the outsourcing of beliefs. Members of different research groups rely on one another epistemically, thereby ensuring that beliefs that are more salient to some members than others are given their appropriate weight and role.

Scientists also outsource beliefs to machines and other tools. As is often the case, this kind of outsourcing is most easily seen when it goes wrong. In 1985, a team from the British Antarctic Survey reported the discovery of a hole in the ozone layer over the Antarctic.[1] It had been known for more than a decade that chlorofluorocarbons (as well as other

[1] I owe my understanding of this episode in recent scientific history to O'Connor & Weatherall (2019) who themselves report a debt to Oreskes & Conway (2011).

industrial chemicals) could deplete ozone. However, the mechanisms underlying ozone depletion had been intensively studied, and the consensus view was that the observations the BAS team reported were impossible. While scientists working on ozone depletion agreed that there was a need to phase out the use of CFCs, they didn't think that there was any urgency. A phased reduction would be sufficient and was well in hand. Hence, the BAS report was met with skepticism.

This skepticism was justified, in large part, on the basis that the theoretical models appeared to show that the rapid depletion required for the BAS report to be accurate was impossible. But it was bolstered by the fact that the BAS data conflicted with another set of data. The BAS team used measurement devices at ground level. A NASA satellite monitoring stratospheric ozone levels had failed to detect any significant change. Not only were there good theoretical reasons to think that the BAS data was wrong, there was also conflicting evidence; the likelihood was that the BAS team's data was due to some failure of measurement or processing.[2]

Nevertheless, Richard Stolarski, a physicist based at the Goddard Space Flight Center, decided to double check the satellite data. He discovered that in fact the satellite *had* detected the anomalously low levels of ozone the BAS team reported. But *because* these measurements were anomalously low, the software had removed them as outliers. This kind of preprocessing of data is a necessity in science, since bad readings and instrument failures are not all that uncommon. Building theoretical presuppositions into our equipment is therefore routine and a requirement of good science. We make implicit assumptions about what entities we will measure, and what the properties of these entities might be, and embody them in our instruments and our algorithms. The scientists had

[2] Anomalies like this are not all that rare in science. It is an extremely difficult question when the best response to an anomaly is to devote time and resources to attempts to resolve it and when it is better to ignore it. In this case, a relatively cheap and rapid review was sufficient to resolve the anomaly. If that investigation had not sufficed, it might have been rational to live with the anomaly, assuming that, given the theoretical models and the satellite data, the BAS measurements must be wrong, even in the absence of a satisfying explanation of how they came to generate their findings. Anomalies and unexplained errors are a fact of life: instruments glitch, people make mistakes in recording values or in adding them, and so on. If we make a mistake in ignoring an anomaly, the future progress of science will usually correct for this, because genuine data that cannot be satisfactorily explained will cause further trouble for our models.

encoded their theories about ozone depletion into their instruments, and because they'd done so, they had failed to detect data that conflicted with these theories.

The story has a happy ending. The theories turned out to be essentially accurate: the problem arose not from the scientists' understanding of how ozone interacted with industrial chemicals, but from a failure to factor in how extreme cold and the Antarctic winds influence the process. It also had a happy ending inasmuch as the problem of ozone depletion was solved by international agreement to phase out CFCs (despite a familiar story of industry resistance, epistemic pollution and US recalcitrance). In any case, it is an excellent illustration of how scientists outsource beliefs to instruments. Many scientists who relied on data from the NASA satellite had no idea how, or even that, the data was corrected. They could outsource beliefs about the need for correction (to account for measurement error) to the software.

Ordinary people outsource belief in both these kinds of ways too. We rely on others to believe things on our behalf, and we outsource our beliefs to the external world. Some of the ways in which we outsource belief are already familiar to epistemologists. Goldberg (2010) for instance has persuasively argued that we rely on others to maintain the stability of our beliefs concerning important events and well-known people. I can be said to know that the Queen is still alive because had she died today, I would have been told by now. But our epistemic dependence on others goes very much further. I rely on others to maintain my attitudes about all kinds of things. I even rely on them to help me know what my principles are. I outsource some of my most important and heartfelt values to the community of knowers. The fact that I outsource in this manner helps to explain how I could rapidly replace one principle with another.

Belief Is Shallow

Heartfelt beliefs (for present purposes) are those I feel strongly about and would defend fervently. They *feel* deep. Some of them *are* deep, in the sense that they're ingrained. They might be ingrained because they cohere with and are supported by other states that are constitutive of me

as a person.[3] Perhaps they're ingrained in a different sense: they're somehow encoded such that it would be difficult to overwrite or even to inhibit them. Many beliefs are surely like that. But many fervently and sincerely expressed beliefs are much shallower, both in the sense that they are in some manner optional for me (my identity—again in a characterological sense—is not threatened by my abandoning them) and in the sense that as a matter of fact I would abandon them easily, perhaps without even noticing I was doing so, and even in the absence of (first-order) evidence against them.

The shallowness of belief is an instance of a more general phenomenon: the adaptive outsourcing of cognition to the world. Evolution is sensitive to small differences in costs, and it is cheap to outsource representations. Why go to all the trouble of building a model of the world, for instance, when the world is easily available to represent itself? As an added bonus, the world is a more accurate representation of itself than any model could be. It represents itself without any loss of detail, on a handy 1:1 scale, and updates in real time. As Rodney Brooks famously put it (1990: 5), "the world is its own best model." So long as the costs of accessing the world are not significantly higher than the costs of accessing an internal model (and they may often be lower, not higher) we ought to expect the job of representing the world to be taken on by the world itself. That is, rather than consult an inner model, we should expect organisms to use sense perception to track the world and how it changes over time.

And that seems to be exactly what we find. Just as we are stubborn epistemic individualists, despite the fact that we defer to others pervasively, so we take ourselves to have rich internal representations, but actually retrieve much of our visual imagery from the world as and when we need it. We are subject to a fridge light illusion: we think our model of the world is rich and detailed because whenever we attend to any aspect of our model, we retrieve a rich representation of that aspect (see Chater (2018) for a lively presentation of the evidence; some of it will already be familiar to philosphers from Dennett (1992)). But our

[3] Here "person" is used in a characterological, not a re-identification, sense; see Schechtman (1996) for this distinction.

internal models are sparse. For example, if a computer monitor is set up to display constantly altering garbage text, but with real words appearing and disappearing so that their appearance is timed to coincide precisely with our saccades as we read, we have the experience of reading an unchanging page (Rayner 1998). We take ourselves to have a rich and stable representation of the page, but we fail to notice how it changes. Change blindness experiments provide further evidence for the surprising poverty of our internal representations. In these experiments, images that are identical but for one (large) feature are presented to participants successively, interspersed with a flicker (to prevent retinal persistence). Even on repeated viewings of the same pair of images, the apparently obvious difference between the images is surprisingly hard to detect (Simons & Levin 1997). Change blindness has been demonstrated in the world outside the laboratory, with passers-by failing to notice when the stranger they were talking to was replaced by another who shared only the broadest similarity to the first (Daniel J. Simons & Levin 1998).

Daniel Simons, one of the researchers responsible for the work on change blindness just mentioned, has also been involved in work on inattentional blindness that further illustrates how sparse our perceptual representations may be in the absence of attention. In a now famous experiment, Simons & Chabris (1999) had participants watch a video of two teams of players passing a basketball. In the easy condition, participants had to count all the passes; in the hard condition they had to keep separate tallies of passes for each team. While the teams were passing the ball, another person walked through the players, (apparently) in full view of the watchers. In one condition, the person wore a gorilla costume; hence the paper is called *Gorillas in our midst*. In the hard condition, only about half the participants noticed the gorilla. Even in the easy condition, around a third failed to notice the gorilla (in both conditions, they were more likely to notice a woman carrying an umbrella than the gorilla, perhaps because such an intrusion is less unexpected; nevertheless, collapsing across all conditions around a third of participants failed to notice the woman). Again, this seems evidence that our visual representations are less detailed than we'd have expected.

These experiments provide persuasive evidence of what we should, on theoretical grounds, have expected in any case: we offload aspects of

cognition onto the world. As Andy Clark, who has done more than anyone else to highlight the extent of our cognitive dependence on the world and on artifacts has argued,[4] we believe in accordance with the 007 principle: keep things on a "need to know" basis:

> evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done. (Clark 1997: 64)

Clark's focus is on how tools and artifacts extend our cognitive capacities. There are, however, good reasons to think that beliefs are extended in analogous ways.

A central piece of evidence that our visual representations are extended consists in their *shallowness*, where a representation is shallow if it is easily uprooted. If beliefs are extended, then we should expect to see evidence that they are shallow in similar ways. We take ourselves to have rich and detailed internal representations of the visual scene, but careful experimentation shows that we actually retrieve a great deal of visual information by rapid and unconscious saccades as and when we need it. As a consequence, our visual representations are shallow: under appropriate conditions, we may fail to notice the substitution of one person for another, because our visual representation is (largely, though of course not entirely) stored outside our heads. Similarly, we take ourselves to have stable beliefs, but if our beliefs are shallow then we should expect them to be easily uprooted through analogous processes.

Cognitive dissonance experiments (reviewed in Cooper 2007) provide experimental evidence that our beliefs can be easily uprooted. In one

[4] Clark's account of cognitive dependence—which pays relatively little attention to social dependence—comes in defense of the extended mind: the thesis that minds are not confined to the skull of agents, but instead leak out into the world (Clark 2003; Clark & Chalmers 1998). The extended mind hypothesis is highly controversial, with a number of philosophers arguing that our cognitive dependence is better explained by our minds being *embedded* rather than *extended* (Rupert 2009). I'm not taking sides on this dispute. Holding that our cognition is heavily dependent on the environment and on other agents doesn't commit me to holding that our minds are literally extended (nor to rejecting that claim).

variant of the induced-compliance paradigm, participants are asked to write essays defending a conclusion that they can be expected to find unpalatable. Often, college students are asked to defend the conclusion that their tuition fees should rise. Participants in the control group are paid to write the essays, while those in the experimental group are prevailed upon to write them (the experimenters might tell the participants that they can argue either for or against tuition rises, but that they don't have enough essays arguing in favor of a rise, and it would be nice of the participant to write such an essay). The manipulation results in those in the experimental group being significantly more likely than those in the control group later to express support for the conclusion they defended in the essay.[5]

Why should the manipulation induce those in the experimental group to be significantly more likely to endorse the conclusion they defended in the essay? The induced-compliance manipulation is designed to leave participants feeling free to refuse to write in favor of tuition rises (course credit or payment, for instance, doesn't depend on what they write). The perception they chose freely to write an essay defending tuition fees puts the participants in a very different situation to those in the paid group: they find it harder to explain their own behavior to *themselves*. Because they can't attribute their choice of topic to the effect of financial inducement or to constraint, they take their actions as evidence that they believe, or at least are not violently opposed to, the conclusion (Carruthers 2013). A simple manipulation of their own behavior is sufficient to induce a shift in their beliefs.

[5] The classic work on cognitive dissonance stems from the bad old days in psychology, when sample sizes were often very small and questionable research practices abounded. For that reason, we shouldn't place too much weight on these findings. A multi-lab preregistered replication (using the essay writing task and an induced compliance manipulation) is nearing publication. Previous preregistered replications of some of the basic findings (e.g., Forstmann & Sagioglou 2020) provides grounds for guarded optimism with regard to the replication. Moreover, the researchers behind the multi-lab replication attempt report that they believe that the effect is real, albeit inflated in the published literature by the file-drawer effect (Vaidis & Sleegers 2018). As we noted in the first chapter, researchers tend to have an accurate sense of which effects are real (Camerer et al. 2018), so their expectation is also (some) grounds for optimism. More recent work on choice blindness—soon to be discussed in the main text—suggests a mechanism for belief revision akin to that at work in these experiments, which is a third reason for optimism, or at any rate for believing that a mechanism like this one helps to cause belief revision.

The ease with which we are led to self-attribute beliefs by these kinds of manipulations suggests that we lack detailed internal representations of our beliefs. Just as our internal representations of the visual scene lack detail, and are easily swamped by changes in the external world so long as gross features are retained (*white man dressed as a construction worker*, as in Simons and Levin 1998), so our beliefs can be swamped by quite weak evidence that we believed something else all along. Notice, though, that this isn't just good evidence that we lack detailed internal representations. It is also good evidence of how we make up for this lack: by reference to cues for belief. In this case, the cues are our own behavior. In other cases (we'll soon see), these cues are the behavior of other people. Under a variety of circumstances, we will tend to construct or reconstruct our beliefs on the spot, rather than recalling them, with attention to cues to belief central to this reconstruction.[6]

We saw above how change blindness experiments provide striking evidence for the sparseness of our perceptual representations. *Choice* blindness experiments provide parallel evidence for the sparseness of our beliefs, and, like the evidence from cognitive dissonance, suggest that beliefs about our own values and commitments are much shallower than we'd have expected. In choice blindness experiments, participants make a series of choices between pairs of options, and their responses are recorded. For instance, the options might be represented by cards and the chosen card placed in a pile. After the participants have completed the series, the experimenters show them the options they picked one by one, and ask them to justify their choices. On a minority of questions, however, the choices have been switched and participants are asked to justify a choice they didn't in fact make. Participants rarely notice the switch (despite being given the opportunity to correct their responses). Instead, on most switched trials the participant goes on to defend choosing an option they had actually rejected. They attribute

[6] Constructing beliefs rather than recalling them is in fact routine. As Gareth Evans (1982) influentially noted, we (typically) don't answer questions like "do you believe there will be another World War?" by searching through a repository of our beliefs, but instead by considering the world. In the kind of case I have in mind just as much as in Evans' case, we construct our beliefs by assessing the evidence for them. Evans focuses on first-order evidence (are geopolitical tensions rising?), but higher-order evidence (what do people like me think? What do *I* think?) is evidence too.

to themselves a belief that is at odds with one they'd expressed just a few minutes earlier.[7]

Choice blindness has been demonstrated in multiple spheres, from judgments of facial attractiveness (Johansson et al. 2005) to moral judgments (Hall et al. 2012). Choice blindness has been demonstrated with regard to political opinions too. In the lead up to a Swedish general election, researchers approached members of the public and asked them about their attitudes to the actual policies of the parties contending in the election (Hall et al. 2013). They were asked to express their attitudes to each policy by marking a point on a 100mm line, with one end indicating complete agreement and the other complete disagreement. They were then asked to justify their choices, including some choices which had been switched (the median alteration changed the point marked by 35.7mm). Only 22 percent of manipulated responses were detected and more than 90 percent of respondents offered at least one justification of an altered choice. Self-reported political engagement and the use of extreme ends of the scale didn't correlate with likelihood of detecting a switch. Similarly, most participants in a study of hot-button moral issues offered justifications of choices they hadn't in fact made (Hall, Johansson, & Strandberg 2012).

In these studies, just as in cognitive dissonance experiments, belief revision was induced by manipulating cues to what the person themselves believes. In others experiments, belief revision is induced by manipulating cues to what *others* believe (taking advantage of our disposition to engage in social referencing). In one experiment, Maoz, Ward, Katz, & Ross (2002) presented Israeli Jews with a real draft peace proposal. Jews who were told that it had been drawn up by Palestinian representatives were significantly less favorable to it than those who were told that it had come from the Israeli side (which was in fact the truth). In a second experiment, Maoz et al. showed that Palestinians too were susceptible to devaluation of a proposal on the basis of information about who authored it. Cohen (2003) reports even stronger evidence: for his sample, information about whether welfare policies were supported by House Democrats or House Republicans was a more powerful

[7] Note that choice blindness has been replicated with a massive sample (Rieznik et al. 2017).

predictor of attitudes to it than policy content. Much more recently, Barber and Pope (2019) showed that endorsement by Trump powerfully influenced Republican attitudes to a policy, independently of its content. Liberal policies supposedly endorsed by Trump were 15 percent more likely to be supported by those on the political right than those which were not (supposedly) endorsed by him (conversely, being told that Trump *opposed* a policy didn't move liberals).

These studies appear to demonstrate that many beliefs are shallow, in much the same kind of way as perceptual representations are shallow. I take myself to have rich internal representations of the world around me, but in fact these representations are quite sparse: the fact that I will easily accept substitution of one obtrusive element in my visual scene for another without noticing is evidence that my state lacks the detail I take it to have. I find it hard to hang on to the visual scene, because in fact my representation of the world is *out there*, in the world: I rely on the stability of the world to ensure my representation doesn't wobble. Similarly, I take myself to have fixed and definitive beliefs. But I rely on the world (where the "world" includes my own body) to tell me what I believe and to ensure both stability and detail in my beliefs. If the world changes, my beliefs change with it, sometimes without my noticing.

Outsourcing and Belief Shift in the Real World

The evidence just cited is evidence that our belief states are often much less rich and much less stable states than we would have guessed. While of course we have internal belief states, they are remarkably shallow. When we ask ourselves what we believe, we look as much to the world (especially the social world) to answer the question as to our own internal states. In the face of evidence that *that I believed that p*, or that *people like me believe p*, I conclude *I believe p*, and I may do so even if previously I had quite a different belief, and without noticing the shift.[8]

[8] No doubt, alternative explanation, consistent with our having rich internal models, might explain the data. But our interpretation of experimental results should be constrained by other data and the theories that explain them. We should expect to see outsourcing to the world, when conditions are appropriate (when, for instance, our perceptual access to the world is

Let's turn from the laboratory to the real world, to see these mechanisms of belief attribution at work. Take the case of Never Trumpers once more. This is a disparate group of individuals, and no doubt different mechanisms are at work in different members of the group. But social referencing might explain belief revision in some of these cases (or it might partially explain it, by explaining how the revised belief came to be more palatable or more accessible). When Never Trumpers made their initial declarations, Trump was one candidate among others for the Republican nomination (Graham's February 2016 proclamation was typical). At that point in the nomination process, social referencing didn't result in a strong signal of support for any one candidate. Instead, insofar as someone who strongly identified with the Republican party asked themselves "what do people like me believe?" the response was divided among the possible candidates. Because the signal was divided, other sources of evidence and motivation had to take up the slack: Republicans had to make up their minds who to support by reference to a familiar array of considerations (electability, personality, policies, personal identification, and so on) with social referencing playing a more minor role. As the primary season progressed, however, candidates fell away one by one, and the signal constituted by the support of other Republicans for the remaining candidates strengthened. The signal supporting Trump in particular may have grown in strength as a consequence of the numbers attending his rallies (and as a consequence of the media's constant highlighting of these numbers).

Once Trump was declared the presumptive nominee in early May, the signal returned by social referencing was unified: *people like me (largely) support Donald Trump*. Given that we discover what we believe, in important part, through such referencing, the signal constituted pressure to fall into line. And as Never Trumpers fell away, the signal increased in strength. Once Trump was elected in November 2016, the signal gained additional strength: *people like me support the President*, especially when he comes from my party.

sufficiently rapid and reliable). We saw in the previous chapter that there is a strong case for thinking our species owes much of its success in colonizing a great variety of environments to cultural evolution, and the way in which it distributes epistemic labor. It is in light, too, of this picture of ourselves that we ought to interpret the evidence.

This kind of mechanism supports rational belief update, moreover. We'll have a lot more to say about the rationality of mechanisms like this one (which I'll argue are common) in later chapters. For the moment, let's just note that this kind of mechanism manifests *ecological* rationality (Todd & Gigerenzer 2012). A cognitive process is ecologically rational when it is designed to get the right answer, no matter how it does it. Mental shortcuts are the paradigm cases of ecologically rational mechanisms: it might not look rational to guess that one foreign city has a larger population than another because the first is more familiar to you, but the recognition heuristic works, at least in some domains (Goldstein & Gigerenzer 2002). Social referencing is rational in this kind of way. Recall from the last chapter that groups of deliberators are (under appropriate conditions) more accurate than individuals. Recall that knowledge is distributed across individuals, and that deference to group norms is adaptive given this fact. Recall, too, the prestige bias: other things equal, it is adaptive to imitate the beliefs (and behavior) of prestigious individuals. Given all these facts, there was *rational* pressure for someone like Graham to update his beliefs. The majority of people like me support Trump; prestigious individuals with whom I identify (or up to whom I look) support Trump; therefore *I* support Trump. In later chapters, I'll argue that social referencing is rational directly: it's not just a way of reaching the right result; it's a way of reaching the right result by responding to evidence.[9]

In many cases, this shift will occur without any phenomenology or other indication of a change of mind. Instead, the person may take themselves to have always believed what they now believe. Our memories are highly fallible. Perhaps this fallibility arises from the same causes that explain the fragility of our beliefs and our perceptual representations: because they're very sparse. There is extensive evidence that memory is not a snapshot of the past, with details preserved, but is instead

[9] The recognition heuristic itself might work in this way and therefore qualify as directly rational. A mechanism is ecologically rational if it is well-designed to generate the right response (in an appropriate environment). It is directly rational if the cues to which it responds are not merely correlated with getting the result; they're *evidence* for it, sufficiently strong to justify the response given. Recognition might be evidence for city size; hence relying on it may be relying on justifying evidence.

sensitive to the circumstances of retrieval. As a consequence, memory exhibits shallowness parallel to that of other representations. Loftus (2003) has shown how vulnerable our memories are to false information, with participants in her experiments incorporating into their memories information that was actually presented after the events. Schacter (1996) cites evidence that this kind of effect is not confined to the lab, but plays a role in the confabulation of sometimes spectacularly false and highly consequential memories: memories of being the victim of abuse, and even of being an abuser, for instance. It would not be entirely surprising, therefore, were some people to shift from fervent opposition to equally fervent support of a politician without taking themselves to have changed their minds.

Of course, a prominent Never Trumper like Graham won't be allowed to forget he was once a fervent opponent of the president. The media and political foes will delight in reminding him. But the sparseness of memory and of representations more generally may nevertheless aid someone like him in reconciling himself to his shift. He may imagine all sorts of mental reservations, for instance. Sparseness leaves room for such confabulation: probing his mental states, he'll fail to find a unified, obdurate, *Never Trump!* representation. Rather, he'll find little content beyond the words themselves, and he'll have room to confabulate all kinds of caveats (*Never Trump*, *unless* he were to become more presidential; *unless* he's endorsed by Ted Cruz, or what have you). For others, who are less likely to be reminded of previous views, the shift comes more easily. Those White Evangelicals who suddenly dropped the view that excellence in character was essential in a politician may well see themselves as holding a single position over time.

It bears repeating that this mechanism of belief update is not partisan. It's one we should expect to see deployed across the political spectrum. People will differ in who they defer to, and what counts as a good cue to belief, but they will engage in analogous kinds of social referencing. Social referencing may help explain relatively rapid shifts in opinion on left and right. Take the relative swiftness with which gay equality has achieved widespread acceptance in most developed nations. It wasn't long ago that gay people were routinely mocked and caricatured on mainstream television. Within my memory, such mocking became unacceptable to most

people, and being gay became relatively unremarkable, as the candidacy of Pete Buttigieg for the Democrat nomination, and his subsequent cabinet selection by Joe Biden, illustrates.[10] Polling data on marriage equality provides quantitative evidence for this decline in prejudice (Pew Research Center 2019). Opposition to marriage equality halved, from 60 percent of US adults in 2004 to 31 percent in 2019, while support doubled over the same time period (31 percent/61 percent). The shift is too big and rapid to be very significantly explained by a cohort effect, and in fact we find shifts in every cohort, including the oldest. Of course, there are multiple explanations for this shift, and (first-order) evidence and argument surely played some role. But social referencing probably had some effect too. Insofar as people settle what they believe by asking themselves *what do people like me believe?*, changes of mind can snowball. The more people shift to believing *p*, for whatever reason, the more *people like me* believe that *p*, for any value of "people like me."[11] Shifts in belief can help to cause further shifts in belief, and we may see very rapid changes in attitudes.

All that said, it's possible that this particular mechanism tends to operate more strongly on the right than the left, today at any rate. It is sometimes held that the contemporary left is characterized by identity politics, rather than by identification with a broader movement. If this is right, or if (for some other reason) identification with the Democratic party or any other broad institution is weaker,[12] we might find that the effects of social referencing are weaker, because for those on the left there are fewer people (sufficiently) *like me* to provide a strong signal for

[10] This is of course not to suggest that anti-gay prejudice is no longer a huge problem. Rather, the point is that it has decreased to a very significant extent in the population, in a relatively brief period of time.

[11] Of course, this is the case only when the issue is not strongly polarized. If only one side of a debate shifts, then asking oneself *what do people like me believe* will yield divergent answers for (say) Republicans and Democrats, and we may see snowballing on one side and intransigence on the other (think, for example, of how attitudes to Russia moved in lockstep on the left and right until recently, when they began to diverge sharply; Suls 2017). While attitudes to same sex marriage remain somewhat polarized, the degree of polarization is relatively small and Pew reports large shifts on both sides of politics.

[12] One piece of evidence in favor of thinking that identification with a single bloc is weaker on the left than the right: media consumption patterns are much broader on the left than the right, with the former accessing both highly partisan media and more mainstream sources, while the latter confines itself much more to partisan sources (Benkler et al. 2018).

belief update, or because signals are spread across multiple groups. If that's the case, of course, it's not because people on the left are more rational or that their beliefs are explained by different kinds of mechanisms. The same mechanisms will operate in different ways and with different effect in different contexts.

The Pervasiveness and Rationality of Outsourced Belief

Ever since Putnam (1975) drew our attention to the phenomenon, it has been a familiar fact that we rely on experts to fix the reference of some of our concepts. I don't need to know how to distinguish elms from beeches because, when relevant, I can exploit the expertise of relevant experts to make the distinction for me. In fact, this kind of deference is very common. My "elm" concept is adaptively indistinct. Rather than having fully specified contents, our representations often include *placeholders* or *pointers*, where a "placeholder" is a gap or indistinctness that requires specification (if it becomes necessary to deploy a specified concept) and a "pointer" is a placeholder that indicates where specification may be sought (Rabb et al. 2019). As Rabb et al. recognize, scientific concepts are typically adaptively indistinct in just this kind of way.

It's not just scientists who defer like this, of course. On the basis of testimony from their parents or teachers, children come to accept many claims that they can't distinctly represent. They may accept that *germs cause diseases* while having little idea what "germs" are, for instance. These indistinct representations may scaffold the later acquisition of more precise first-order representations (Sperber 1996), but many of us never come to replace our indistinct first-order representations with anything more precise. Most of us have little idea what "viruses" are, even while the world is gripped by a pandemic we know to be caused by one. Many people believe *that the theory of evolution is true*, but experimental evidence indicates that many of those who take themselves to "believe in" evolution have little genuine understanding of the theory or the mechanisms involved (Shtulman 2006). This indistinctness is not a bug, but a feature: given that genuinely understanding the science is extremely difficult (such that even experts can master only some of the

relevant knowledge), our having indistinct concepts allows us to defer to the experts to provide specification. But experts engage in this deference too: theories outside their area of expertise constrain their work, sometimes in ways that require them to allow others to provide details in their own representations.

Indistinctness of content is also common in the domain of political and moral beliefs. One reason to think that our political commitments are indistinct is that most of us sign up to political programs that are amalgams of several conflicting currents, apparently without really noticing the conflicts. For instance, mainstream conservatives support political parties that are strongly in favor of free markets and of traditional institutions like the family. But as we previously noted, free markets are solvents that undermine traditional institutions. Mainstream center-left parties often combine irreconcilable currents within them too (a commitment to the environment and to traditional, highly polluting, blue collar jobs, for instance). Even at an individual level, there is great indistinctness of content. As Schwitzgebel (2009b) notes, concepts like "freedom" and "brotherhood" and propositions like "all men are created equal" are widely affirmed, but are "only half-filled or quarter-filled with a real thought" (57). They are grasped indistinctly, and await outsourcing to others to get a distinct content.

This outsourcing leaves us vulnerable to sudden shifts, of the sort seen in the case of Evangelical Christians or Never Trumpers. Our disposition to outsource in this kind of way constitutes a vulnerability, because once others know what cues we respond to, they can seek to exploit us by exploiting these cues (more on this in later chapters). But *vulnerability* doesn't entail *defect*. Being vulnerable to others in this kind of way may be adaptive. There are no free lunches, and every adaptation has costs. Our vulnerability may be an inevitable cost we pay as epistemically social animals, and being epistemically social is responsible for much of our capacity to generate significant knowledge. Outsourcing is ecologically rational. It is also *directly* rational: being open to cues for what others believe is being open to *reasons*. A proper defense of this claim must await later chapters. For now, let me emphasize how recognizing the rationality of these kinds of dispositions should change our attitudes to our fellow citizens.

The media delights in telling us how ignorant we are. Often this takes a partisan form (we've all seen the videos of Trump supporters making outrageous claims; we rarely pause to wonder how many people they had to ask to get the response they wanted), but it also occurs in a non-partisan, if equally self-congratulatory, way. The surveys reported are often badly designed—or well designed to get the kind of answers that will generate headlines—but the underlying finding is robust: most of us know surprisingly little about history, geography, or contemporary politics. Some people know a lot, but "most people know nothing, and many people know *less* than nothing" (Brennan 2016: 24). Brennan (2016) surveys some of the lowlights of this ignorance: even during election years, most people cannot identify the candidates standing in their district; most cannot name the party that controls congress; they routinely report the opinion that the US spends too much on foreign aid, instead suggesting that a percentage of the budget that vastly *exceeds* actual spending would be appropriate; 40 percent can't name the United States' enemies during World War II; at the height of the cold war, only a minority correctly stated that the Soviet Union was not a member of NATO. More than a third of actual voters are "know-nothings" Ilya Somin (2013) claims. Such ignorance of basic facts concerning the issues central to elections is by no means confined to the United States. A survey of people who voted in the United Kingdom's Brexit referendum revealed that neither Leave nor Remain voters knew much about the EU; on some questions they performed *worse* than chance (Carl et al. 2019).

Voters' ignorance concerning the issues at stake in elections is central to Brennan's (2016) case for epistocracy: government by the informed. It is easy to sympathize, in the light of election results in Brazil, the United States, India, Turkey, and elsewhere over the past five years: isn't this what government chosen by the ignorant looks like? But it would be a mistake to blame the apparent rise in populism on voter ignorance alone: voters have *never* been well-informed about the issues or the parties; instead they make their decisions by looking to signals from elites (J. R. Zaller 1992). That's not irrational: that's adaptive outsourcing of belief. Voters may not know how their representatives promote their values, but they know who to look to for such promotion. They may not know what policies the parties offer, but they often have "metaknowledge"

(Rabb et al. 2019); pointers to repositories of information that fill in the gap. Even if they lack such metaknowledge (having mere "placeholders"), they can outsource cognition by deferring to experts who have more detailed knowledge (sometimes, deference may appropriately be to those with metaknowledge, rather than those who know).[13]

Possession of metaknowledge isn't necessarily inferior to possession of knowledge. We live in a complex world, and we all daily engage with systems and use concepts we cannot fully specify. None of us is expert in anything more than a small range of domains. Metaknowledge unaccompanied by knowledge may not merely compensate for our deficits; it may often be *better* than knowledge. While the genuine expert may be cognitively better equipped than the person who has metaknowledge without knowledge, the person with metaknowledge alone may be better equipped than the person who has *some* knowledge, but falls short of expertise. The person with metaknowledge may track changes in expert opinion much more reliably than the person with knowledge short of expertise, since the former defers to the experts for the contents of their beliefs.

It's in this light that we should interpret our use of concepts like "freedom" or "equality," when they're quarter-filled with a real thought. It's intrinsically difficult to fill them with a thought—that is, to specify what they mean. Philosophers who specialize in these topics disagree with one another, and would probably agree that their preferred view has costs, inasmuch as it fails to capture dimensions that others emphasize. Only specialists can hope to make these concepts anywhere near fully distinct. Yet we non-specialists may *appropriately* be committed to them. Outsourcing specification to the experts may be the rational way for us to proceed. It's not a limitation on our rationality; it's a way in which we manifest it. In a slogan, meta is (often) better: outside the very limited sphere in which we can acquire detailed knowledge, we often do better and cognize more rationally by knowing how to find out whether *p*, or knowing who knows whether *p*, rather than by knowing

[13] Again, this outsourcing leaves us open to exploitation by those who can make use of it, of course. The degree to which we're vulnerable to exploitation is limited by the fact that we remain sensitive to first-order evidence regarding how well we're doing: voters turn on the governing party when economies falter, for instance. We integrate our higher-order evidence with the first-order evidence that is sufficiently near and clear for us to make use of it.

much about *p*. Sometimes, we even do fine by deferring in the absence of metaknowledge.[14] We are all very ignorant, and we should be fine with that.

Of course, there are kinds of ignorance that really are costly. Our ignorance of our pervasive reliance on the outsourcing of cognition and on metaknowledge might be a costly kind of ignorance. In fact, we seem to mistake our outsourced knowledge for individually possessed knowledge and think of distributed cognition as inferior. Paradoxically, our sense of how much we know *individually* is sensitive to our metaknowledge: in response to cues that others know, and cues that we can access testimony, we inflate our estimate of how much we know *without* the need to rely on others (Rabb et al. 2019). Access to the internet increases people's confidence that they can answer questions *without* using the internet; the knowledge that one's partner in an experiment knows something inflates a person's confidence that *they* know it; people rate *their* sense of understanding higher when they are led to believe that experts understand something compared to when they are not (importantly, however, this effect was seen only when they were told that this knowledge could be accessed), and so on. We are subject to a pervasive *illusion of explanatory depth* (Mills & Keil 2004; Rozenblit & Keil 2002), arising from a similar mechanism. The illusion of explanatory depth refers to our tendency to overestimate the extent to which we understand and can explain the workings of everyday objects and natural phenomena. People express confidence in their ability to explain how flush toilets and piano keys work, or what causes tides and rainbows, which greatly outstrips their actual ability to explain these things. Keil & Wilson (2000) suggest that this illusion arises from the pervasiveness of the division of epistemic labor and our effortless facility with it. So ingrained is deference and so fluent that we confuse knowledge possessed by our community with knowledge possessed by us.[15] At the

[14] Consider how we defer to the geographical knowledge of taxi drivers or pilots. We may lack metaknowledge in some such cases: we may not even know *who* is in charge.

[15] I once suggested that conspiratorial ideation may arise, in part, from the same mechanism: our vulnerability to mistaking community knowledge for individual knowledge leaves us with an inflated sense of our capacity to explain complex events; in some individuals, this leads to suspicion of a coverup when the official narrative does not strike us as fully satisfying (Levy 2007).

conclusion of this chapter, we'll return to this kind of ignorance. While first-order ignorance may not be a problem, higher-order ignorance—ignorance, perhaps, of the very fact that first-order ignorance is not a problem—may play a role in our current epistemic malaise.

How to Make Up Your Mind

I've suggested that the pervasive outsourcing of belief should lead us to rethink our attitude to the apparent ignorance of our fellow citizens. The most consequential example of such apparent ignorance is, of course, ignorance of climate change. Understanding the mechanisms underlying belief acquisition and update should lead us to rethink ignorance of climate change too. Contrary to a widely accepted story, it doesn't arise from stupidity or irrationality. It's a product of mechanisms responding to evidence. Understanding these mechanisms will provide us with levers for changing people's beliefs, including their beliefs about climate science.

Why do so many people reject the science of climate change? Because the social mechanisms of belief update provide them with good reasons to do so. Centrally, they deploy social referencing, asking *what do people like me believe*? Multiple cues tell them that people like them reject the science (think of how advertisers and merchants of doubt play on cues to identity and identification: true Americans drink Coke; true Americans are individualists who won't be told what to think by outsiders). Kahan thinks of mechanisms like this as identity-protective and philosophers have followed his lead. For instance, Carter and McKenna (2020) argue that we have strong grounds to be skeptical about beliefs acquired via these mechanisms, because they are responsive to "the epistemically irrelevant property of being socially approved." But that's a mistake: the fact that a proposition is socially approved is higher-order evidence that bears on its truth, and there's nothing irrational in being guided by it. The primary purpose for which we deploy these mechanisms is to get things *right*, not (just) to fit in.[16] So too with other mechanisms

[16] In defence of a view somewhat similar to Kahan's, Williams (2021) suggests that an epistemic account like mine cannot explain why we see worse performance from those who score

that seem partisan—for instance, giving more credence to testimony from those who are like me than those who are not. There is experimental evidence that we give more weight to the explicit testimony of those we see as benevolent toward us (P. Harris 2012; Mascaro & Sperber 2009; Sperber et al. 2010). This bias is rational because those who don't share my values may seek to exploit me, and those on my side are likely to be more trustworthy (toward me). This kind of mechanism may generalize beyond explicit testimony to the implicit testimony constituted, for example, by *failing* to attend to certain messages (thereby signaling they're not worth attending to), or simply by getting on with one's life without worrying about the climate. If those I identify with don't worry about climate change, that's a reason—a genuine *reason*—for me not to worry. They provide me with evidence.

Asking oneself *what do people like me believe* is one mechanism whereby people come to reject (or to accept) climate science. There are, of course, others, and they work in concert for many skeptics. They also ask *what do prestigious people* believe? Different people accord different degrees of prestige to different individuals. For liberals, the relevant individuals may be scientists (think of the role Dr. Fauci played in the pandemic), as well as politicians on their side. Conservatives get a very different message about climate change from opinion leaders on their side: most notoriously from Donald Trump, but also from most other major figures in the Republican party, as well as from the apparent experts they see on Fox.

Evidence of consensus is significant for both sides. Of course, there's a consensus *among scientists* about climate change, but the consensus that is most weighty for an individual might be the apparent consensus that is reinforced every day: the apparent agreement of those around them. Who we interact with day to day plays an important role in what we believe. Bob Inglis, a former Republican congressman who lost the backing of his party and his seat over his endorsement of climate science, credits his conversion from skepticism to belief in climate change

higher on tests of knowledge and capacity. One possible explanation is that greater sophistication allows for better deference: those who know more know better what they ought to believe in virtue of their ideology. See Martin and Desmond (2010) for a suggestion along these lines.

partly to visits to US bases in the Antarctic, where scientists showed him the evidence (Cohn 2013). There's every reason to think that the (first-order) evidence played an important role, but living in close proximity to the scientists and thereby coming to establish a basic level of trust with them likely also played a role. We pay attention to cues for consensus (P. Harris 2012), but since our best evidence for consensuality will often be what those around us say, those nearest to us will once again have an outsize role in what we take to be consensual (mere repetition can induce the impression that something is widely believed; Weaver et al., 2007).[17]

The environment also contains multiple non-verbal cues to belief; cues that are also genuine evidence. It's hard to sustain belief that we face a crisis when those around us are relaxed. It's hard to sustain the belief that we must dramatically reduce our consumption of fossil fuels when the gas station is central to your daily life, or when your job or those of your friends depends directly or indirectly on oil. It's also hard to sustain belief in a climate crisis when it is cold outside, and a number of studies have found that belief in climate change increases during hot spells and falls during cold, even among those who distinguish between weather and climate (Borick & Rabe 2017; Hornsey et al. 2016). Perhaps the sheer solidity and robustness of the built and natural environment makes it hard for us to believe we face a crisis (perhaps the coronavirus pandemic might unsettle our taken-for-granted assurance that the world will simply continue as before sufficiently to make us more receptive to worries about the bigger challenge represented by climate change).

Bad belief in climate science has—formally—the same source as better belief. Those who accept the science see scientists as prestigious and the institutions that employ them as reliable. Those who do not have much less positive views, both of scientists (trust in science has been

[17] The most effective content-based intervention for producing better beliefs on climate change seems to turn on making perceptions of the scientific consensus more accurate (since a consensus of genuine experts is strong higher-order evidence, that's not surprising). Perception of the consensus is a "gateway" to more accurate belief (S. L. van der Linden et al. 2015, 2019). But it's not sufficient simply to present that information: in the face of multiple dissenting voices, informing people about the true degree of consensus doesn't correct people's misperceptions. It's necessary to "inoculate" people by first presenting information rebutting—or prebutting—denialist talking points (S. van der Linden et al. 2017).

declining on the right since the 1970s; Gauchat 2012) and of universities (58 percent of people who are Republican or Republican-leaning now say that colleges and universities have an overall negative effect on the United States, compared to just 19 percent of those who are Democratic or Democratic leaning; Pew Research Center 2017). We pay attention to those we trust, those whose values we share (or who we take to share our values), those who are prestigious or simply near to us, and we believe what they believe. Doing so is rational: it's a mechanism for calibrating belief to evidence, albeit indirectly. Like all mechanisms, however, it can misfire, and in the case of climate change it has gone spectacularly wrong for very many people.

If you want to change minds—your own or those of others—you need to look to the social and institutional cues on which beliefs depend. Change those, in the right ways, and minds follow. If social referencing yields a particular answer to the question what do *people like me believe about X*, then I'll tend to accept that answer as my own. If cues to belief are distributed across the social and institutional environment in certain ways, then those who are sensitive to cues like these will tend to accept the associated beliefs. Change minds by changing the world: physical and social. That, I suggest, is not only how minds are most effectively changed. It is how minds have *always* been changed, at least with regard to matters outside the immediate purview of the agent.

Outsourcing Religion

I'll finish this chapter with a brief digression from the main argument, to address a question that was posed, but not answered, in Chapter 2. In that chapter, we briefly surveyed evidence of behavioral inconsistency among religious believers. Christians appear to be more likely to bid in a charitable auction on a Sunday than on other days of the week (Malhotra 2010); consume less pornography on a Sunday than other days (Edelman 2009) and think of the dead as living on only when given a religious prompt (P. Harris & Giménez 2005). Christians are by no means unique in this kind of apparent inconsistency, which is also seen in Hindus (Xygalatas 2012), Muslims (Duhaime 2015) and in the Vezo people of

Madagascar, who practice an ancestral religion (Astuti & Harris 2008). Van Leeuwen (forthcoming, 2014) thinks these inconsistencies are best explained by differences in the kind of attitude people take to religious claims. In response, I pointed to evidence that religious credences tend to govern behavior more broadly than Van Leeuwen thinks, and that therefore they qualify as beliefs. But if behavorial inconsistency is *not* explained by the nature of the underlying attitudes, what does explain it? I suggest that it might arise from the outsourcing of cognition.

We outsource not only to people and cognitive artifacts, like books and the internet, but also to the environment: for example, we outsource perceptual representations to the world around us. We surely outsource religious representations to other people: *I* don't need to understand the nature of the Trinity, because I can rely on theologians to do so on my behalf. But perhaps we outsource religious representations to features of the environment too. Whereas when I outsource my visual representations, the world takes on the role of (partially) constituting the representation for me, outsourcing of religious representations is more likely to play the role of providing prompts for more purely internal representations (call such triggers *cues*). It's adaptive to outsource in this way, because it increases the efficiency with which cognitive resources can be managed. If I can be confident that my representations will reliably be cued when they're needed, I don't need to be vigilant for situations to which they're relevant. This kind of outsourcing can explain inconsistencies in behavior across time: in the absence of cues, my behavior won't be guided by a representation that would otherwise be relevant to me.

Religious representations can be expected to differ in the degree to which they're cue-dependent and the range of cues to which they're sensitive. They might also differ in how long they persist once triggered, and these differences may themselves be a function of how reliably they're cued. Consider the apparent difference in persistence of the Sunday effect (the dispositions of Christians to behave consistently with their religious creed to a greater extent on Sundays than other days of the weeks) compared to the effects of the call to prayer on charitable giving. The call to prayer appears to have short-lived effects on behavior, with a rapid drop off in its effects on charitable giving (Duhaime 2015).

A mere 20 minutes after it ended, the percentage of those choosing the most charitable option had dropped from 100 percent (while the call was audible) to less than 50 percent. In striking contrast, the Sunday effect appears to persist long after Church services have ended.

It's unlikely, I think, that this difference represents a deep difference in Islam versus Christianity. More plausibly, it is the result of a difference between the availability of cues to religion in a very secular society, like the contemporary United States, versus a more religious society like Morocco. In the contemporary US, reminders of religion are much less frequent than in Marrakesh. Because reminders of religion are less frequent, believers need to internalize them to a far greater degree: once they're cued (for instance, by the realization that today's the Lord's Day), they persist longer and decay slower than elsewhere. For those who live within the hearing of a mosque, by contrast, the call to prayer is available five times a day, every day, to take up the representational slack. There's much less need to internalize the representations. In medieval Europe, I speculate, church steeples and bells might have played an analogous role, and the Christians of Paris or Prague may have had fast-decaying religious representations, like contemporary stallholders in the Souks. Conversely, Muslims in Melbourne or Manitoba would have persistent religious representations, more like their Christian counterparts than their co-religionists in other countries.

4

Dare to Think?

In the last two chapters, I argued that beliefs are pervasively outsourced to other people and to the environment, and that belief revision often occurs in response to changes in the cues that scaffold our beliefs. I've suggested that in light of these facts, we need to ensure the scaffolding of better beliefs. We need, that is, to manage the epistemic environment. Bad beliefs are produced by a faulty environment and better beliefs are best promoted by environmental engineering.

This kind of suggestion is certain to face opposition. Resistance will come on at least two fronts. First, changing minds by changing the environment—rather than by giving people *reasons* to change their minds—seems manipulative. I'll address that concern in Chapter 6 (where we'll see that the objection is mistaken, not because it's not manipulative to change minds without giving reasons, but because changing the environment in the relevant ways *is* giving reasons). In this chapter and the next, I'll focus on the prospects for cognition in the *absence* of environmental engineering and other kinds of scaffolding. I'll argue that individual cognition (unaided) is very much less powerful than we tend to think. Worse, if we refuse to aid the individual cognizer, others will be all too eager to fill the gap, often in ways that leave them, and us, worse off.

The view I defend in the next three chapters has a passing resemblance to a more familiar position, one I used to endorse myself. This position is also sometimes offered in response to worries about epistemic paternalism, and holds that Enlightenment ideals of intellectual autonomy are psychologically unrealistic; instead paternalistic measures must take up the slack (Conly 2013; Levy 2012; Moles 2007). This familiar response appeals to the rationality deficit view of ourselves we outlined in Chapter 1, according to which contemporary psychology has

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0004

shown that rationality is a scarce resource and that for the most part we respond relatively unthinkingly. That's not a view I now endorse: one of the aims of this book is to argue that we are in fact rational animals. If epistemic engineering is justified, it's not because we're not rational enough to respond to epistemic challenges. It's because we're never sufficiently rational on our own.

Regulative Epistemology

Of course, it's easy to find examples of individual reasoning going wrong. Too easy. Piling up example after example won't go far toward showing that individual cognition isn't powerful (we can also, it should be conceded, find plenty of examples of individual reasoning going well). We need to be fair to individual reasoning: we want to focus on examples when it is conducted carefully and thoughtfully. I'll focus here on what might reasonably be taken to be individual reasoning given its best shot. I'll focus on individual reasoning in the form recommended by leading contemporary epistemologists who explicitly aim to develop practical guides to good reasoning. I'll focus, that is, on reasoning as recommended by regulative epistemology.

Traditional epistemology is concerned with the nature of knowledge and of related concepts, like justification. Traditional epistemology is in one sense rather *un*traditional: the narrowing of focus to these questions represents a departure from an earlier way of thinking about knowledge that was more explicitly regulative. Philosophers like Locke and Descartes were concerned with how we can think better—with how we can *attain* knowledge—as well as with the nature of knowledge. The tradition to which Gettier and his many intellectual offspring belong represents a newly narrowed focus for epistemology (see Ballantyne 2019 for a sketch of this history).

Regulative epistemology offers us news we can use: practical precepts and advice on how to think better, in the pursuit of knowledge. As it has been developed to date, regulative epistemology is individualistic. It is addressed to individual thinkers, offering them advice on how to gather and collate evidence, how to weigh it, how to avoid error and what warning

signs on the path to knowledge to look out for. Regulative epistemology has been developed most fully by virtue—and vice—epistemologists. While the distinction between regulative and analytic epistemology was introduced by Nicholas Wolterstorff (1996), its currency in debates today is due principally to Roberts and Woods, who subtitle their book *Intellectual Virtues* "an essay on regulative epistemology," and it's to virtue and vice epistemology that I will turn first. Virtue epistemology, in its regulative form, may have a role to play in guiding us toward better belief, but I will suggest that its role is extremely limited.

Virtue epistemology, at least in the "responsibilist" variety that will be our focus here, recenters epistemology around virtues and vices.[1] Virtues and vices are character traits (and, perhaps, ways of thinking and attitudes as well) which are, respectively, epistemically helpful and harmful. Either directly (by making us more or less responsive to evidence or to criticism, for example) or indirectly (by making us love truth or to be indifferent to it, to read widely or to be incurious, and so on) they help or harm our functioning as epistemic agents, especially the extent to which we're able to acquire knowledge. In its regulative guise, virtue epistemology might offer guidance in one or more of several broad ways.

Virtue epistemology in its regulative guide might adopt what Roberts and Wood call "the social engineering model of the regulative philosopher's role" (Roberts and Wood 2007: 29) and aim at the inculcation of virtues in us, for example, by advising us on how we should educate our children or ourselves, or more directly, by trying to shape the virtues of its readers. Alternatively or in addition, the regulative epistemologist might aim to influence our enquiry prior to (perhaps as a means toward) our becoming virtuous. She might urge us to conduct our enquiry in the way a virtuous person would conduct enquiry, or to believe what we

[1] It is this ("responsibilist," rather than reliabilist) variety of virtue epistemology that has sometimes turned regulative. Whereas virtue reliabilism thinks of virtues as reliable cognitive faculties (e.g., perception or memory), responsibilist virtue epistemology mirrors virtue ethics in thinking of the virtues principally as character traits, like generosity or humility. Both varieties of virtue epistemology were initially developed as attempts to address the nest of issues that have occupied epistemologists since Gettier upset the game board in 1963; principally, offering an analysis of knowledge. Reliabilist virtue epistemology's principal architect is Ernie Sosa (see the essays collected in Sosa 1991); Linda Zagzebski played a similar role in the development of responsibilist virtue epistemology (Zagzebski 1996).

would believe were we virtuous, or something along these lines. No matter how it goes about it, I doubt that virtue epistemology (in its regulative form) will be especially helpful to us.

I won't discuss any of these different models for regulative epistemology in any detail. Whatever their value in other ways (perhaps the inculcation of the virtues is intrinsically valuable, though I'm somewhat skeptical),[2] I'll suggest that the virtues are not especially valuable as a means of regulating our epistemic conduct in the service of the acquisition of knowledge. Perhaps the virtues conduce to knowledge, but they do so to a very limited extent, and then only in appropriately structured epistemic environments. Inculcation of the virtues is helpful, if it is helpful at all, not as an *alternative* to scaffolding inquiry and structuring the environment, but only in *conjunction with* these measures.

Inculcating the Virtues

Virtue epistemology, in its regulative guise, aims to improve belief by inculcating the virtues in believers; correlatively, it explains some or many cases of bad belief by reference to a lack of intellectual virtues (or the presence of intellectual vices). Virtue epistemologists offer analyses of specific virtues and guidance on how they manifest, with the aim of allowing us to identify them in others and develop them in ourselves. The intellectual virtues are the cognitive parallels of the ethical virtues. Just as we ought to strive to be and to act out of virtues like generosity, courage and honesty, so we ought to strive to be open-minded, epistemically humble, charitable and so on, and conduct our thinking in a way that manifests these virtues.

Virtue theories, in their ethical and their epistemic guises, face the charge that they fail to be action- (or conduct-) guiding. These high-minded phrases are at best difficult to pin down. It's difficult to give

[2] Elsewhere, Mark Alfano and I have expressed doubts that possession of the epistemic virtues conduces to the generation of knowledge. We argue that much of our most significant knowledge is generated collectively, by agents who individually manifest epistemic vices, rather than virtues (Levy & Alfano 2019).

anything like a precise content to the virtues, at least in a way that's principled and doesn't beg crucial questions. Famously, Aristotle identified the virtues as a mean between two extremes: courage is the mean between cowardice, on the one hand, and foolhardiness, on the other. In doing so, he highlighted the problem: identifying the virtues requires us to engage in interpretative activity and interpretations are easy to contest and difficult to defend. If I'm accused of cowardice in not standing up to that bully, I may retort that that would have been foolhardy, not brave. How can we settle this dispute in a principled way? Virtue theorists often appeal to practical wisdom at this point; that is, to a kind of intellectual capacity to judge cases well. At best, that's little use to those of us who apparently lack practical wisdom. How are *we* to proceed, now? At worst, it might be nothing more than an appeal to the inherent superiority of an interlocutor to put an end to a dispute.

Virtue epistemology faces the same set of problems: the intellectual virtues, too, are at best hard to pin down in a principled way. Virtue epistemologists (and virtue ethicists) have a reasonable response to this worry, however. They may dismiss the wish for anything more precise to guide action and intellectual conduct as a fantasy, insisting that doing and thinking the right thing really just *is* a matter of difficult judgment. They'll point out, moreover, that they don't simply appeal to practical wisdom to end debates. They offer detailed arguments and analyses of the virtues, arguments, and analyses that can bring those of us who currently lack practical wisdom a little closer to understanding the virtues and a little nearer to practical wisdom ourselves.

I'm agnostic on the extent to which virtue epistemology and its analyses are genuinely helpful. If virtue epistemology can help, however, it's not by substituting for apt deference to others and socially distributed cognition; instead, it's by playing a (small) role in helping us to do these things better. Virtue epistemologists may be on to something, but in its current guise their (explicit) recommendations are far too individualistic. Explicitly at least, they appear to aim to bring us each to inculcate the virtues in ourselves and then, guided by our intellectual excellences, to tackle hard problems largely on our own. *That* strategy must fail; or so I'll argue (in fact, many of the examples of intellectual excellence they

themselves cite involve agents whose epistemic success is due mainly to apt deference, and not—much, at any rate—to the deployment of the virtues).

I'll work through a case study of sorts: I'll focus on one virtue epistemologist's attempt to delineate the virtue of open-mindedness from a kind of intellectual "flaccidity," on the one hand, and from dogmatism, on the other, and to provide concrete guidance for intellectual inquiry.[3] I'll suggest that the enterprise is misguided as a central focus of an explanation of bad belief and a remedy for it. Even if we succeed in delineating open-mindedness and inculcating it in ourselves, we won't have the tools we need we for better beliefs. For that, we need other people and effective epistemic institutions much more than we need virtues.

Open-Mindedness as an Epistemic Virtue

The vulnerability of open-mindedness to excess is a part of folk wisdom. We all recognize that there's such a thing as being *too* open-minded. "Be open-minded, but not so open-minded your brains fall out," as the quip has it. We don't think that the virtuous agent will give just as much weight to the health advice of their physician on the one hand and their cousin who posts memes about vaccines on Facebook on the other. They shouldn't be equally open-minded about each.

Perhaps we shouldn't be open-minded at all about those vaccine memes; perhaps we should just dismiss them more or less unthinkingly. Kripke (2011) argues that dogmatism—a paradigmatic epistemic vice—is warranted in the face of certain claims. He gives the example of astrology: "I once read part of a piece by a reasonably well-known person defending astrology […] I was not in a position to refute specific claims but assumed that this piece was of no value" (Kripke 2011: 49). Rather

[3] Admittedly, Cassam, my stalking horse, is a self-described vice theorist, not a virtue theorist. Nevertheless, he seems an apt theorist to discuss under the heading of virtue epistemology. Vice epistemology follows the example of virtue theory in explaining bad belief as the upshot of the character traits and dispositions of agents (Cassam adds "attitudes" to the mental states that constitute vices, but he thinks that virtues too are sometimes constituted by attitudes). Moreover, the epistemic vices are the mirror images of the epistemic virtues: they imply one another and each can be analyzed in terms of an extreme lack of the other.

than being open-minded with regard to every claim we encounter, he suggests, we should "delineate cases when the dogmatic attitude is justified" (Kripke 2011: 49).

As Cassam points out, dogmatism looks very much like a manifestation of the "archetypical epistemic vice": closed-mindedness (Cassam 2018: 39). Kripke advocates refusing to consider evidence and arguments against conclusions he takes to be settled. Cassam argues that such dogmatism is epistemically vicious, and advocates serious engagement, even under conditions like these. Moreover, he suggests, dogmatism threatens knowledge. Kripke had argued that dogmatism can protect knowledge: when we know that an argument leads to a false conclusion, but can't see how to rebut it, we do best to refuse to engage (see Fantl 2018; Levy 2006 for related arguments). Cassam argues that on the contrary, dogmatism undermines knowledge, on the grounds that knowledge requires justified confidence in a belief. If we don't have a right to our confidence, because we maintain our belief dogmatically, we don't genuinely *know* the proposition we believe.

Of course, Cassam accepts that the virtuous agent should be slow to abandon their justified convictions in the face of arguments they can't immediately see how to refute. We should never be dogmatic, he maintains, but we often ought to be appropriately firm in our opinions (open-mindedness is the mean between intellectual flaccidity and dogmatism). Kuhn (1970) influentially argued that scientists are and should be much less open-minded than we tend to think. Scientists are very reluctant to abandon their "paradigms;" the set of findings, exemplars of good scientific practice, methods, and assumptions that for them constitute good science. In the face of an observation that conflicts with a well-established theory, they'll usually reject the observation, rather than the theory (as we saw in the previous chapter). That looks like dogmatism: rejecting a claim or an observation as false, simply on the grounds that it conflicts with what we expected to see.

Cassam accepts that it's entirely appropriate for the scientist to assume that the anomaly can be accommodated by the paradigm (or, alternatively, that it's the upshot of measurement error), even if she can't immediately see how. That's not dogmatism; that's appropriate "firmness or tenacity" (113). What distinguishes firmness from dogmatism is that

whereas the dogmatic scientist would shrug her shoulders and move on in the face of the anomalous, the firm scientist will work to accommodate it or to show that it's spurious. The virtuous scientist is not dogmatic, because she's willing "to acknowledge fundamental flaws in established tools and beliefs, and abandon those tools and beliefs" (113).

Cassam thereby offers us a genuine alternative to Kripke, not a mere relabeling of what the latter calls "dogmatism" as "firmness." Kripke argues that in the face of an argument for a view we know to be false, sometimes we should just shrug and move on. Cassam advocates engagement with these arguments. If we refuse to engage, he maintains, we lose our right to confidence in our beliefs and thereby lose knowledge. One should be confident in one's beliefs, and that requires confidence in our capacities to tackle spurious arguments. I'll argue that insofar as Cassam urges us—ordinary agents, who lack any special expertise in the domain of the argument—to tackle these arguments on our own, he's wrong: engaging with them risks knowledge to a far greater extent than does dogmatism. Reason, argument, evidence, and practical wisdom unaided leave even the fully virtuous person vulnerable. Without heavy-duty social and environmental scaffolding, even virtuous agents can't reliably acquire knowledge about difficult and complex issues. The right response to spurious arguments is often to shrug and move on, relying on others to tackle them for us.

The dispute between views like Cassam's and those more like Kripke's turns in important part on an empirical question. Kripke asserts, and Cassam denies, that it is sometimes very difficult to discover where spurious arguments go wrong; difficult enough that we are more likely to lose knowledge by confronting a spurious argument than by dogmatically refusing engagement. Kripke maintains he couldn't see where the argument in favor of astrology went wrong and implies he was unlikely to be able to identify any flaws quickly had he persisted. Cassam thinks this shows a vicious lack of intellectual self-trust on Kripke's part. *Of course* he could identify flaws: the more dubious a view, the easier it is to debunk. When we're confronted with dubious claims—like our cousin's anti-vaxxer rants on Facebook—we must respond seriously, or we risk

our knowledge. We must identify the flaws, do our own research and (when technical expertise is necessary) consult the experts.

I'll argue that Cassam is wrong on the empirical question: in fact, we're at much greater risk of losing knowledge from "doing our own research" than we are from dogmatism. It *is* true that we are often able to rebut spurious claims, but that's not by probing them for ourselves: it's by apt deference. The intellectual virtues play only the smallest of roles in any of this. I'll also argue that Cassam is wrong in his characterization of the behavior of scientists. Scientists often are (appropriately) untroubled by the anomalous. They don't work to accommodate it. They shrug and go on, dogmatically. If that's how scientists (our paradigms of epistemic excellence) should behave, it's even more the case for the ordinary person. We, too, should often shrug and be dogmatic.

Climate Change Skepticism, Holocaust Denial, and Other Fantasies

One of Cassam's principal aims, in developing vice epistemology, is to enable us to understand the origins and persistence of conspiracy theories and the like.[4] He argues that epistemic vice is an important factor in explaining why people accept these theories. If he's right, then we should find that the epistemic virtues are protective against them: the virtues enable us to see through these theories. Moreover, he's committed to a further claim: there won't just be a correlation between possession of the epistemic virtues and the rejection of conspiracy theories and the like. There'll be a direct causal connection: agents will reject such theories *because* they're virtuous and as a result of *deploying* these virtues in

[4] David Coady (2007) and Charles Pigden (2007) both urge that we drop the pejorative use of "conspiracy theories," because conspiracies are all too often real and we risk unjustifiably stigmatizing conspiratorial explanations by this usage. I have some sympathy for this view. I use the term "conspiracy theory" here somewhat tentatively, to pick out that subset of theories that postulate a conspiratorial explanation of events where that explanation runs counter to the explanation offered by duly constituted epistemic authorities (see Levy 2007 for discussion). Given our pervasive dependence on testimony, rejecting such explanations is usually irrational, even if the explanation is in fact false.

assessing these theories.[5] Cassam is committed to thinking that the epistemic virtues will make a significant difference in our capacity to assess such theories accurately, because they'll enable us to see through them.

We're each exposed to a massive amount of misinformation; many of us daily. Our cousin may post his anti-vaxx memes on our Facebook wall; we may see "plandemic" fliers on the street and the news may report the false claims of certain politicians. If Cassam is right, at least some of this material requires a serious response from each of us: we are able to retain knowledge *that COVID-19 is not a hoax* only if we each confront the claims of those who maintain that it is. How hard can it be?

Many of these claims venture into areas of specialist knowledge. I don't have specialist training in immunology or in climate science. That puts me at an immediate disadvantage when it comes to assessing these claims, especially when they stem (directly or indirectly) from those who do possess specialist training in these areas. I take myself to be reasonably sophisticated and reasonably knowledgeable about climate science, which I've been following for more than a decade. But very often, when I come across sophisticated denialism I find myself in the same position that Kripke took himself to be in reading that article on astrology: I know it's wrong, but I have no real idea how.

Here's an example I happened to come across recently. Visiting the website for Springer publishing (a reputable publisher, for whom I edited a journal for a decade), I was presented with an advert for a book called *The Rise and Fall of the Carbon Dioxide Theory of Climate Change*. The author is one Rex Fleming, who has a PhD in atmospheric science from the University of Michigan and is a fellow of the American Association

[5] Mark Alfano and colleagues have found that the virtue of epistemic humility correlates with lower rates of acceptance of conspiracy theories and lower rates of acceptance of fake news (Meyer & Alfano forthcoming). While this is evidence in favor of Cassam's view, it's correlational: it doesn't show that agents reject conspiracy theories due to the deployment of epistemic virtues. As a matter of fact, I'm skeptical of the causal claim. Meyer and Alfano used a sub-scale from a previously validated epistemic humility scale (Alfano et al. 2017) to measure the virtue. I suspect that this sub-scale is transparent to respondents. The items (sample items: *I don't take people seriously if they're very different from me*—reverse scored, of course—and *I appreciate being corrected when I make a mistake*) are clearly designed to probe dispositions that we, as good liberals, are *supposed* to value. People respond accordingly. Both rejection of conspiracy theorizing and the scale tap into the same underlying dispositions: dispositions to respond in the way that left-leaning, educated Western individuals value. I suspect sociological factors explain the correlation, not the virtues.

for the Advancement of Science. Fleming has had a decades' long career as a scientist and has published work on climate science and on modeling in reputable journals, including the *Journal of the Atmospheric Sciences* (impact factor 3.194) and *Environmental Earth Sciences* (impact factor 2.18). Fleming is also a climate change skeptic: he argues that CO_2 has no impact on global warming.

If Cassam is right, we—you as much as I—now have an obligation to confront Fleming's claims seriously, if we're to retain the knowledge *that climate change is real and CO_2 is a very major contributor to it*. We can't just shrug our shoulders and move on. According to his own website, Fleming's book establishes:

> the failure of the Schwarzschild radiation integrations to maintain the CO_2 longwave radiation intensity achieved in the surface warming by H_2O and CO_2. The resultant Planck radiation intensity is severely depleted in the upper atmosphere. The result is the CO_2 molecules merely pass their remaining small residual heat to space un-impeded.

To retain our knowledge, we're going to have to rebut this claim, show it's irrelevant or that it's gobbledygook. If we can't, we may lose our right to confidence.

Can you do this? I have to confess that I can't (not in the way that Cassam recommends: by deploying my intellectual capacities to grapple with the first-order evidence). I'm at an immediate disadvantage when it comes to the task: I have little idea what the words I just quoted actually *mean*. I'm tempted to dismiss them as gobbledygook (a commentator on the website *Skeptical Science* dismisses Fleming as a purveyor of half-baked physics he learned on Wikipedia), but his track record of publication in reputable scientific journals and his other credentials makes me suspect that something more sophisticated than the mere repetition of ill-understood verbiage is going on. I'm going to have to do better.

How should I go about it? I can turn to Google and search for discussions of Fleming's book (that's how I came across the quote from *Skeptical Science*), but most of this discussion is beyond me. Perhaps Wikipedia would serve me better: there is, I see, an entry for "Schwarzschild's equation for radiative transfer," but even the Wikipedia

entry is very heavy going for me. It's math-heavy, and I'm extremely limited in that area. Even coming to understand what Fleming is claiming—rather than getting on with assessing his claims—is likely to be the work of many hours for me. Without the math, I doubt I'd *ever* be in a good epistemic position to genuinely assess and rebut Fleming. Even with it, I'd likely require hundreds of hours of immersion in the technical literature to rebut the claims of someone with a PhD in the relevant area and a decent track record of publication on technical questions like mathematical modeling. Since much of that literature is currently beyond me, I'll have to start small: maybe with a high-school math textbook (and I'd better get going on the syllabuses in chemistry and physics too). I'm going to have to acquire genuine expertise to rebut Fleming's claims (at least, that is, to rebut them in the kind of way Cassam wants: by identifying for myself the flaws in the arguments).

How much genuine expertise am I going to need? I don't know enough to know how much I don't know or how much ignorance I need to remedy. Would a good undergraduate degree in climate science suffice? Maybe, although the fact that Fleming publishes in the area makes me think that may not be enough. Of course, Fleming is far from the only sophisticated skeptic; the famed 97 percent consensus on climate change is, if anything, an underestimate of the proportion of those with relevant expertise who accept the mainstream view on climate change, but that still leaves room for hundreds of dissenters. If the issues are difficult enough that the experts sometimes get them wrong, what chance do *I* have, working without institutional support and colleagues and acquiring the needed skills along the way?

Let me illustrate the difficulty with a different example, from an area in which the issues are surely less complex. In *Vices of the Mind*, Cassam discusses implicit bias at some length. He concludes his survey with the observation that such biases are malleable and respond to efforts on the part of motivated individuals to change them (173). I happen to possess some degree of genuine expertise on this topic, on which I've published multiple papers (see Levy 2017, 2016, 2015, 2014a, 2014b). I've read many papers on the very topic Cassam here considers: the extent to which our biases respond to efforts at self-cultivation. Yet I remain unsure whether Cassam is right in claiming that they're malleable in the

way he suggests. The literature on implicit biases is large, but it's tiny compared to the multi-disciplinary literature on climate change. The range of expertise required for it is correlatively small. If I haven't been able to answer the implicit bias question for myself, I despair at my capacity to rebut sophisticated climate skeptics.

It's worth adding that the multi-disciplinary nature of climate science (like many other areas of contemporary science) entails that many actual climate scientists may lack the skills to rebut the sophisticated skeptic. Only a minority of climate scientists work on issues like radiative transfer; the rest work elsewhere, on different issues. Some (in my ignorance I have no idea how many) of those who work on different issues will have a working knowledge of radiative transfer, and perhaps that'll enable them to rebut Fleming or rapidly to acquire the capacity to rebut him. Few will waste their time in this way. Few climate scientists want to take time off from their research to answer the arguments of a crank: if those they trust tell them the points raised have been dealt with, they'll shrug and move on. Surely that's all we can reasonably ask of them. If I'm right that the majority of climate scientists would have to take a significant amount of time away from their research to rebut Fleming for themselves (how much time? In some cases, mere hours; in others, weeks or months), then we can't expect each of them to invest this time. They'd have no time for their own research if they had to do this with regard to every crank they encountered.

It may be important that *someone* rebuts the claim (that will depend on its novelty). The great majority of climate scientists will outsource the job, and defer to whoever does it. For the most part, this deference will itself be dogmatic: they won't search for rebuttals. Rather, they'll move on, confident that if the claims are worth engaging with, someone well-placed to do so will take on the task. They'll deal with challenges they take to be worth taking seriously in their own areas instead. They know if the challenge that Fleming raises is genuinely troubling for the field, the news will reach them (it would reach all of us soon enough). In the meantime, dogmatism is justified epistemically and pragmatically.

Climate science is highly unusual, and perhaps taking it as our case study unfairly tilts the odds against Cassam. After all, climate science is highly specialized and highly technical. Many readers, like me, will face

a high barrier to entry into discussions into the area due to its reliance on mathematical modeling and very specialist knowledge. But it's surely not out of place to focus on the single most significant case of bad belief in the contemporary world: if virtue epistemology struggles to handle this case, what's it good for? Nevertheless, perhaps it has a purchase in less technical areas. Cassam's own principal example in *Vices of the Mind* is Holocaust denial. Perhaps on a topic like this, virtue epistemology can make significant inroads, even if it struggles with climate science.

It is surely true that the barriers to entry into discussions of history are lower (at least for likely readers of this book) than into climate science. In academic history, claims are conveyed largely in natural language, not the technical languages of mathematics and statistics, and we're all experts at parsing natural language. We should, however, take care to refrain from the philistinism of those who regard only the sciences as genuinely worthwhile intellectual pursuits. History isn't just story-telling. It has its own tools and techniques and its own experts. We're not experts in history or in adjudicating historical claims merely in virtue of being expert language users. To see through the lies of sophisticated Holocaust deniers like David Irving (on whom Cassam focuses), we need field-specific expertise, where the field is not simply "history" but more precisely twentieth-century German history (or more precisely still, World War Two history or even Holocaust history). Perhaps we can acquire that expertise more rapidly than we can expertise in radiative transfer, but it would nevertheless be a very significant investment of time and energy for any of us.

An incident that occurred just a few months after *Vices of the Mind* was published illustrates just how demanding and specialized historical expertise really is. In May 2019, Naomi Wolf gave a series of interviews to promote her new book *Outrages* (Wolf 2019). This book argues that the persecution of gay men in Britain increased dramatically in the second half of the nineteenth century. Wolf pointed to the occurrence of the term "death recorded" in the records of the trials of men accused of "sodomy." She argued that this indicated an increase in the use of capital punishment by the courts and a correspondingly harsher attitude. She suffered the on-air embarrassment of having her error pointed out by Matthew Sweet, a radio host and cultural historian. As he pointed out,

"death recorded" was used for a nominal death sentence: when the accused was found guilty of a capital offence but not executed.

Wolf's public embarrassment was widely taken to indicate that she hadn't done her research diligently and hadn't used fact checkers. While I'm in no position to assess the quality of her research, prima facie she had every reason to be confident in her claims. The book was based on her 2015 University of Oxford doctoral thesis: it had, therefore, been supervised by an expert (in nineteenth-century literature) and had been examined by independent experts. She had also employed a genuinely expert fact-checker to verify her interpretation of the law. Dame Helena Kennedy, a prominent human rights lawyer, had interpreted "death recorded" in the same way Wolf had (H. Kennedy 2019). Wolf's DPhil and her book were both checked by multiple experts, but that wasn't enough to allow her to avoid embarrassing error. What's needed, in science and the humanities alike, is *specific* expertise: expertise not just on that period, but on that practice, in that place.

Our genuine expertise as natural language users may lead us to miss this fact, and to think we can wade in for ourselves, in the humanities if not in the sciences. But we're liable to be mistaken, and embarrassments like Wolf's may easily result. Wolf took care to have her claims checked; others get themselves into trouble because they approach history casually, certain that their native intelligence and common sense equips them to adjudicate historical claims. 2019 seems to have been a treasure trove of such incidents: in that year the journalist Cokie Roberts took it upon herself to charge historians writing about abortion with distorting the record (Wulf 2019). The historians she targeted had claimed that advertisements for abortion services were plentiful in nineteenth-century newspapers. Roberts denied this, on the grounds she couldn't find any. What was missing wasn't the ads, which were indeed plentiful, but the capacity to identify them by the euphemisms they used. Being a good journalist didn't equip Roberts to call out the historians.

If Roberts' embarrassment indicates that historical expertise is needed to adjudicate historical claims, Wolf's indicates just how specific that expertise needs to be. It's not enough to possess expertise in the period or in the law: one needs expertise in the law of that period. Expertise is *brittle* (Kilov forthcoming): an expert in a particular domain is often

unable to transfer their skills to another, intuitively similar domain. Kilov provides multiple illustrations. Surgical skills not only fail to transfer across surgical procedures: they don't even correlate well with performance on tasks designed to mimic them. Performance on surgical simulations is predicted by practice on the simulator, not the possession of skill at the procedure simulated. Professional-level skill in hitting a baseball doesn't equip batters to hit a softball thrown by an expert. Not only do skills not transfer across intuitively similar domains, they also degrade rapidly in response to changes within the domain of expertise: expert bridge players and accountants perform badly in response to arbitrary changes within their domains; the excellent memory for board configurations and moves of expert chess players vanishes when the configurations or moves are not meaningful to them. Expertise in a very specific domain may provide someone with the *confidence* they'll perform well in an adjacent area, but they may nevertheless lack the *competence*.

Of course, expertise can transfer to some degree.[6] The genuine expert often has an edge over a novice, even when there is some degree of mismatch between the domain of expertise and the domain in which it is applied. A historian of, say, modern Europe is surely better placed than I am to adjudicate between David Irving and Richard J. Evans, whose book *Telling Lies about Hitler* Cassam praises as an antidote to Irving. The historian's expertise surely gives her an edge when it comes to evaluating the arguments and evidence each presents. Often, however, expertise fails to transfer: it's far from obvious that my philosophical expertise qualifies me to adjudicate the Evans/Irving dispute any better than an engineer, a plumber or a tax accountant could. Like Fleming, Irving possesses genuine expertise: his 1964 book on the German V-weapons program continues to be well-regarded. This expertise gives him a capacity to fool the naïve—when they attempt to adjudicate claims for

[6] Episodes like those involving Naomi Wolf and Cokie Roberts are examples of what Nathan Ballantyne (2019) calls "epistemic trespassing," where someone with genuine expertise in one field takes themselves to have sufficient expertise to engage seriously with another. Ballantyne has a somewhat more optimistic take on the transfer of expertise than I do. However, his examples of successful transfer involve relatively simple tasks in an expert domain, and the bar for success was set low: expertise transferred just in case the expert performed better than the novice (rather than performed well). Even by this undemanding standard, expertise is surprisingly brittle.

themselves—and ensures that his historical fantasies require genuine and specialized expertise to rebut.

Of course, Cassam recognizes that we often can't be expected to acquire specialist expertise for ourselves, at least when that expertise is scientific (it's much less clear that he recognizes how demanding it would be to acquire historical expertise). 9/11 conspiracy theories often turn on claims about the melting point of steel beams and the like; Cassam recognizes it would be "unreasonable" (117) to expect ordinary people to acquire the knowledge of physics or engineering required to rebut these theories. Instead, he advocates that we consult the experts and work out which of the competing views is correct. While we may not (always) be able to do our own research, in the sense of grappling directly with the first-order evidence, we may deploy the virtues to choose between competing experts.

Cassam is surely right that something along these lines is and ought to be how we should adjudicate disputes like that between Irving and Evans. But insofar as his prescription requires directly and virtuously adjudicating the *second-order evidence*—the evidence that bears not on the truth of the Holocaust, but rather on who is more reliable on this question—it's still too demanding and too individualistic. We face the same risks of losing knowledge by engaging at this level as we do at the first, and dogmatism remains a better strategy. Dogmatism, here, involves the proper scaffolding of inquiry: relatively unquestioning deference to authoritative sources *because they're authoritative* and not because we've assessed their degree of expertise ourselves. It's because they have the right credentials—primarily because they represent the expert consensus view or are endorsed by duly constituted epistemic authorities—that we should defer, not because we've virtuously probed their track records or their citation indexes, let alone because we've evaluated their arguments.

Cassam holds that when we read Evans' *Telling Lies about Hitler*, we'll see Irving's deceptions "brilliantly exposed" (114). But why should we believe Evans' claims? We're in no position to verify his claims about inconsistency with the historical record for ourselves. The same sort of problem arises with other sources we might check. Cassam advises us to turn to Wikipedia and to Google. As he notes, we'll soon learn that the

British High Court ruled against Irving when he brought suit against the historian Deborah Lipstadt for calling him a Holocaust denier, and that he is widely regarded as discredited. But why should we believe the Wikipedia article, or (if we do believe it) that the High Court reached the right verdict? More googling will turn up all too many other sites that support Irving, laud him as a hero, accuse the historical establishment of a conspiracy against him and cite evidence to back up all these claims. How are we to adjudicate any of this?

As a matter of fact, most of us will trust Evans against Irving, and give little weight to Irving's many supporters (some of whom may turn out to have PhDs, alas). We may take ourselves to be convinced by Evans' arguments, judged on their own merits; alternatively, we may take Evans' word for it, but do so because we've diligently assessed his degree of expertise and judged it greater than Irving's. Perhaps, but there's room for a great deal of self-deception here. Why do *I* find Evans' argument more plausible than Irving's? Is it really because Evans' arguments are better and I'm well placed to recognize this? Or am I swayed by him, rather than Irving, because I'm disposed to accept the consensus view? I strongly suspect that's an important element, for me and for Cassam. No doubt Evans' arguments matter, but I bet I'm more receptive to them because I know them to reflect the consensus view, and I suspect the same is true of Cassam. We accept the claims made by Evans and the judgment of the British High Court in very important part *because these are authoritative sources*, and not because they're claims we're in a good position to assess on their merits, or even because we're in a good position to assess how good they are as sources (*that*, too, is a specialist topic, one on which neither he nor I possesses sufficient expertise for confident and well-justified judgment).

Let's conclude this section by turning to the behavior of working scientists, our paradigm of epistemic success. Cassam, recall, denied Kuhn's suggestion that the responsible scientist was a dogmatist who set aside anomalies rather than attempt to explain them or explain them away. In fact ethnographic and historical studies of scientists indicate that Kuhn was right. The scientist often just shrugs in the face of the anomalous, setting anomalies to one side and trusting in the march of science to accommodate them (or to conclude that they must reflect error, since

they aren't replicated). Consider the response of Darwin and his supporters to the work of Lord Kelvin on the age of the Earth and the Sun. Kelvin's genuinely expert estimates of the age of each were vastly too short for the evolutionary account of the origin of life to be plausible. Darwin and his supporters recognized this fact, and recognized that they were unable to refute Kelvin. But rather than following Cassam's advice, and standing ready "to acknowledge fundamental flaws in established tools and beliefs, and abandon those tools and beliefs" (113) in the face of anomalies, they simply set them aside and continued with their work. They relied on the march of science and the work of others, in other fields, to vindicate them, which it duly did later (C. Lewis 2002).[7]

Darwin and his followers were right to set aside the problem, because their research program was so successful in its own domain. This was true even though they recognized that Kelvin's work produced genuine evidence conflicting with their program. Of course, Cassam is right in thinking this strategy has its risks: sometimes the anomalous really is an indication that the research program is flawed. The dispute over the cause of stomach ulcers is an excellent recent example. The primary cause was long believed to be stomach acid, often linked to stress. Doctors were sufficiently convinced of the truth of this theory that some supported fines for those who advocated a rival, bacterial, theory (Zollman 2010). But the bacterial theory was in fact true and the dissidents were eventually vindicated. This little tale doesn't have the moral that Cassam might hope to draw from it, though. It doesn't show that we shouldn't be dogmatic in the face of anomalies. Anomalies are cheap and plentiful. The scientist can't abandon her research whenever she hears of one; that would mean abandoning her research forever. If she is to hang on to her knowledge, she'd better be able to respond by shrugging

[7] In *The Knowledge Machine* (2020), Michael Strevens recounts some of the history of the dispute between Kelvin and Darwin's defenders (Strevens focuses on Kelvin's direct evidence concerning the age of the Earth; Kelvin also argued that the Sun was too young for evolution to be plausible). This episode in the history of science is one of many that Strevens details in which both sides dogmatically (though reasonably) stuck to their guns in the face of evidence they couldn't satisfactorily explain. As Strevens puts it, "Science is driven onward by arguments between people who have made up their minds and want to convert or at least to confute their rivals. Opinion that runs hot-blooded ahead of established fact is the life force of scientific inquiry" (79).

her shoulders and setting aside the many contrarian views she hears expressed every day.

If these reflections are on the right track, virtue epistemology goes wrong to the extent to which it suggests that we can and must secure knowledge by individual cognition. Of course, individual cognition *is* powerful and we ought to deploy it, to the extent we can, in our own fields; perhaps the virtues are an important element in doing so. No matter how virtuously we conduct enquiry, however, we can't rely on it alone reliably to sort out fact from fiction and from fake; even in our own fields, we remain heavily dependent on scaffolding of all kinds for epistemic success. I see no reason to think that the argument won't generalize well beyond virtue epistemology. We non-experts can't hope to rebut the climate skeptics or the Holocaust deniers for ourselves; not without becoming experts (and that requires a great deal of time and effort). Even genuine experts often can't rebut frauds and cranks, if they lack field- and topic-specific specialized knowledge, and those who possess the precise knowledge they need themselves owe their epistemic success to a great deal of scaffolding. By ourselves, none of us are all that epistemically impressive.

Dissent in a Time of COVID

Let me finish this chapter by discussing an apparent counterexample. As I write, the world is in the grip of the COVID-19 pandemic. Governments in most countries claim that their responses are "led by the science" (Peck 2020). Yet very many people, scientists or not, have felt qualified to question that response. Suddenly, everyone is apparently an epidemiologist. Of course, the mere fact that people with no prior expertise feel able to make confident pronouncements at odds with those of (apparent) experts is by itself no surprise. There's nothing new in that. But there may be good grounds for thinking that the pandemic warrants a less deferential response than other areas in which scientific advice guides policy.

Right now, it's not obvious (to me) what the right response to the pandemic is. Most governments have opted for strict lockdowns and strong social distancing measures to reduce transmission, with the aim either

of "flattening the curve"—ensuring that intensive care units are not overwhelmed—or actually eliminating the virus from the population. No one denies that these measures will save the lives of people who would otherwise have died from the virus. Some, however, worry that the economic and social costs of the measures might be greater than the benefits. These are not costs to be weighed *against* the costs to health and well-being: they are *also* such costs. The recession that has resulted from the shutdown of much of the economy across large parts of the world will itself be deadly. The recession that followed the 2008 financial crash is estimated to have led to at least 10,000 extra suicides in Europe and North America (Reeves et al. 2014) and more than a quarter of a million extra cancer-related deaths in OECD countries (Maruthappu et al. 2016). Feelings of isolation linked to the lockdown imposed in many countries will also take a toll on mental health (S. K. Brooks et al. 2020). The economic impact, and therefore (in all probability) the impact on mortality and morbidity of the COVID recession is likely to be much greater this time round. On this kind of basis, it has been suggested that the current response might be more costly than can be justified (Ioannidis 2020). Instead, critics argue, we should collect more data before we settle on a response.

Of course, governments can't wait for more data to come in when they confront a crisis. In this kind of situation, it's widely held they should err on the side of caution, which has been interpreted as entailing that they should lock down hard now, in advance of the evidence. Again, it's not clear to me right now that's right, because it is not clear to me which side is the side of caution. Assessing that issue requires an assessment not just of the health effects of the pandemic, but also of the shutdown. The modeling on which government policy has been based makes assumptions (about infection fatality rates, for instance) which are evolving as we learn more, and has paid little attention to the health effects of the shutdown.

It may be that the actual, highly restrictive, response is partly due to what we might call the goalkeeper's fallacy. There's evidence that penalty kicks aimed down the middle of the goal are less likely to be saved than those aimed to the left or to the right (Chiappori et al. 2002). Part of the reason for this is that goalkeepers usually dive to the left or right to

attempt to save the penalty. They don't stay upright, because they believe (possibly rightly) that they will be blamed less if they made a spectacular and demanding, if futile, attempt to save the penalty than if they engage in the less spectacular strategy of guarding the center. They have an incentive to dive, even if diving is less successful, on average, than not doing so. Similarly, governments may have an incentive to engage in spectacular interventions in the face of a public health crisis. The penalty, in terms of public opprobrium, for underreacting might be much greater than the penalty for overrreacting.

Again, I'm not taking sides here on what the right response should be. I don't know. But even expressing doubts in this way seems quite different to the kind of response I've urged we take to science and scientists generally. What happened to the deference I've argued we should display?

I don't think the pandemic is a counterexample to my claims. The first thing to note is that in expressing doubts about the appropriateness of governments' responses, I'm not expressing any doubts about *epidemiology*. Identifying the appropriate response to the pandemic is not a matter for epidemiologists alone: rather, it's a policy question, on which multiple different kinds of expertise bears. Epidemiologists are not experts in economics or in mental health or social policy or politics or behavioral science, and all these disciplines—and more—are relevant to the right response. To the extent that governmental policy is guided predominantly by epidemiology, it's permissible to worry that it reflects only some of the relevant expertise. Second, much of the debate concerns modeling, and modeling is not the province of any one discipline. People with a range of backgrounds may be qualified to weigh in on models, and those with yet other backgrounds may be qualified to weigh in on their assumptions.

Of course parallel points hold true for climate science as well. The expertise of many different disciplines (economics, sociology, anthropology, political science, even psychology) is relevant to how we should respond to the climate crisis. Why should we be more deferential with regard to climate science than with regard to the pandemic? The central reason is that climate science is mature in a way that our thinking about *this* pandemic is not. This virus differs—in infection fatality rate, in the kind of burden it places on healthcare systems, in the profile of those

who are especially vulnerable to it—from other viruses, and the context of the pandemic is dramatically different from the context of previous outbreaks. Previous pandemics occurred in a world that was vastly different from ours. Think, for instance, of the ways in which the internet has made lockdowns much easier to implement than previously: many people can work or study from home, we can order food to our door, we can entertain ourselves and reach out to others, all very much more easily than ever before. So there's no body of evidence and expertise ready-made to bear directly on *this* pandemic. In contrast, climate science presents us with a consensus which has *already* been tested and retested multiple times, for several decades. A great variety of experts from a great variety of disciplines have *already* contributed to the climate consensus. It's not because the coronavirus is different from climate science that it's appropriate for people to second guess the science. It's because it's the same: there was a time when such second-guessing was appropriate for climate science too. That time has long passed.

There's an important lesson here. A scientific consensus is reliable when it has been stress-tested, by all the disciplines relevant to the topic, for an extended period of time. Only under these conditions is the consensus reliable. Any consensus on the pandemic doesn't meet these conditions. As Schliesser and Winsberg (2020) put it, "there is currently no well-ordered scientific community studying COVID-19 and its impact, so the emerging consensus could be the result of any number of all-too-human biases." These differences between climate science and the state of knowledge over COVID-19 make an epistemic difference: there's no properly generated consensus to defer to in the latter case.

All that said, I'm skeptical that the pandemic is a case in which any of us does better epistemically by making up our own minds. Individual cognition is limited and biased, for reasons that are by now familiar. At this point in the development of knowledge, we may appropriately contribute to the establishment of a consensus through stress-testing, but for each of us it's very unlikely that our considered view is better than that of the epidemiologists advising governments (say). Even in this case, and in the absence of a justified consensus, almost all of us probably do better by deferring than by dissenting—though here the state of knowledge as a whole may benefit if we dissent.

5
Epistemic Pollution

In the last chapter, I argued that individual cognition, even the careful ("virtuous") cognition of those who possess genuine expertise, is a lot less powerful than we tend to think. In fact, it may actually be worse off than the forgoing discussion suggested. I've been arguing that knowledge is dependent on the social and institutional context in which beliefs are acquired and transmitted. But I've paid little attention to the properties of the actual contexts we find ourselves in. We live in *epistemically polluted* environments: deliberately and inadvertently, other agents shape our environments in ways that leave individual cognition even worse off than it might have been. In this chapter, I'll sketch some of the pollutants and how they work to undermine virtuous cognition. The epistemic world has been allowed to degrade, I'll suggest, because we've been unaware of how crucial it is to rational thought. Just as we urgently need to repair and to manage our natural environment, I'll argue, we must repair our epistemic environment.

My focus will be on the so-called "novice-expert problem"; the problem of identifying a genuine or reliable expert among those taking conflicting stances on an issue within their sphere of (apparent) expertise. As Cassam recognizes, solving this problem is essential for laypeople if they are to be able to come to justified views on many important issues. I'll argue that this isn't a problem we should expect novices to solve when they live in epistemically polluted environments.

Novices and Experts

A number of philosophers have risen to the challenge of identifying criteria that ordinary people might use to distinguish reliable experts from unreliable (E. Anderson 2011; Blancke et al. 2017; Johnny Brennan

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0005

forthcoming; Goldman 2001; Guerrero 2017). While there are important differences between them, they converge in identifying *credentials*, *track record*, *argumentative capacity*, *agreement with the consensus*, and *intellectual honesty* as criteria by reference to which we can choose between experts.

Genuine experts have good *credentials*. They have PhDs in the topic under discussion or in a closely related field. They have published peer-reviewed research in the field. Experts with an especially high degree of credibility set the agenda for their field, as reflected in their citation count, and are honored by their peers (E. Anderson 2011). They also have good *track records*, where "track record" consists in more than peer-reviewed publications. It also consists in a record of making predictions that have been borne out by events. Whereas scientific expertise is esoteric knowledge, whether predictions about future events come to pass is often publicly observable and therefore exoteric knowledge (Guerrero 2017).

Argumentative capacity consists in more than debating skill (which can dissociate from genuine expertise). Rather, genuine experts display what Goldman (2001) calls "dialectical superiority." One expert displays dialectical superiority over another if the first expert is able to rebut the claims and arguments of the second. *Intellectual honesty* is displayed by making data available to other researchers, retracting claims that have been refuted and declaring conflicts of interest; because people may be biased, we should heavily discount those experts who have an interest in the truth of their claims. Finally, an expert should be accorded greater credibility to the extent to which her claims are accepted by a *consensus* of her peers.[1]

[1] Goldman (2001) has influentially argued that consensus may not be a good guide to credibility, because the different sources for a claim may not be sufficiently independent of one another. A *non-discriminating reflector* holds whatever opinion their "guru" holds, regardless of its plausibility, and therefore their agreement adds no independent epistemic weight to the initial opinion. In the actual world, agents are never or almost never non-discriminating reflectors. Even young children filter claims for plausibility, and will reject testimony from a familiar person, even a parent, in favor of more plausible testimony from an unfamiliar informant (see Harris 2012). The degree of independence of individual informants from one another varies from case to case, but we can be confident that each filters testimony for plausibility to *some* degree. Of course, experts may nevertheless defer excessively, without being genuinely non-discriminating. Coady (2006) argues that this kind of excessive deference is rare.

While they all recognize that there are obstacles to utilizing these heuristics, the philosophers who have identified these markers of expertise largely accept that ordinary people are able to deploy them to identify genuine experts. I think they're far too optimistic. Ordinary people are well aware that these criteria pick out markers of expertise. But they're also well aware that we live in epistemically polluted environments, and that a major source of epistemic pollution consists in the mimicry of these markers, to inflate the *appearance* of expertise (Guerrero 2017). Our epistemic landscape is polluted, because the cues for expertise don't correlate well with its actual possession. This fact greatly reduces ordinary people's capacity to distinguish reliable from unreliable sources. At the same time, the fact that such deception is widely known to occur reduces trust in legitimate sources.[2]

Merchants of Doubt (Oreskes & Conway 2011) describes a cavalcade of examples of the mimicry of expertise in the service of science denial. Beginning in the 1950s, the tobacco industry responded to compelling evidence that smoking caused cancer by attempting to sow doubt on the science. It aimed not to present an alternative case or to refute the mounting evidence, but to leave ordinary people confused about who and what was reliable. As the infamous 1969 industry memo mentioned in Chapter 1 put it, "Doubt is our product, since it is the best means of competing with the 'body of fact' that exists in the mind of the general public." As Oreskes and Conway document, the tactics the tobacco industry pioneered subsequently spread to those who sought to cast

[2] I suspect markers of expertise are actually more useful to experts themselves than to laypeople. As we've seen, scientists are routinely in the position of needing to rely on the work of other scientists, without being able to assess their work for themselves. They are, however, able to utilize markers of expertise better than laypeople, because they know which journals are predatory, how to assess citations and h-indexes, the quality of particular departments, and so on. In saying this, I take issue to some degree with recent claims by Konrad Talmont-Kominski (2020). Talmont-Kominski argues that in science, the role of source vigilance is very much attenuated compared to other domains. That's not because scientists don't take anything on trust—quite the opposite. Rather, it's because trust is high, Talmont-Kominski suggests. I think the picture is somewhat more complicated. It might be true that the closer the claim is to the scientists' very specific area of expertise, the smaller the role of source vigilance. Nevertheless, source vigilance remains important to scientists. Talmont-Kominski points to the use of double-anonymous review in science as evidence of the bracketing of source vigilance. But that's a revealing mistake: in fact, science uses single-anonymous review much more often than many other areas of inquiry, and its use is often justified on the basis of explicit appeal to the need to know the source of a claim (Palus 2015; Walker & Rocha da Silva 2015).

doubt on ozone depletion, the viability of Reagan's "Star Wars" project and—most perniciously and harmfully—climate change. Indeed, these tactics have spread much further than that. It's not only the well-funded and coordinated industry groups that are the focus of *Merchant of Doubt* that pump out epistemic pollutants. So too do cranks and frauds across the ideological spectrum, whether to convince us to buy their jade eggs or healing crystals, or to believe their theory debunking Einstein.

Charlatans and their fellow travelers employ a variety of tactics to mimic credibility.[3] For instance, those with an interest in deceiving the general public may set up parallel institutions that seem to guarantee expertise. Oreskes and Conway recount how denialists set up the "Nongovernmental International Panel on Climate Change" to peddle myths about global warming. The NIPCC produced reports in identical formats, with identical sections, to those of the IPCC, with the aim of spreading doubt. The American College of Pediatricians utilized similar tactics. The ACP was set up by a small number of right-wing pediatricians to promote their views. Doing so is surely permissible: what's less permissible (and probably intended) was the effect of muddying debate by misleading people into thinking that the college spoke for the profession. When the ACP issued a statement condemning gender reassignment surgery, many people mistook the statement for the consensus view of pediatricians. But the peak body for US pediatricians, the American Academy of Pediatrics, has a much more positive view of gender reassignment surgery (LaCapria 2016). Insofar as the larger organization, with a broader membership base, can be expected to reflect a broader range of expert views and a higher degree of expertise, it is reasonable to give its views greater weight than those of the smaller organization. When the ACP allows or encourages the impression that it speaks for the profession, it introduces an epistemic pollutant.

Fake and dubious journals are also epistemic polluters. The recent growth in predatory publishers, who publish low-quality scientific research for a fee, has attracted a great deal of attention. But the

[3] These fellow travelers, as I label them here, are not seeking to deceive us. They may not even seek to *mimic* markers of credibility. Rather, they may see themselves as creating parallel (but genuine) epistemic institutions and outlets. Nevertheless, their efforts result in the introduction of pollutants into the epistemic environment.

phenomenon isn't new. *Merchants of Doubt* provides several examples: for instance, the climate denialist *Journal of Physicians and Surgeons*. This journal, and its predecessor the *Medical Sentinel*, also published articles questioning the link between HIV and AIDS, the consensus on DDT and papers alleging that abortions are much riskier than the consensus view maintains. For an even more egregious example, consider the case of pharmaceutical companies cooperating—conspiring?—with the publishing giant Elsevier to produce promotional materials designed to mimic peer-reviewed journals (Grant 2009). These fake journals were able to leverage the prestige of Elsevier to give the "research" they published an air of reliability. When the deceit was uncovered, however, the effect was just the reverse: rather than make the findings published look more legitimate, the deception made Elsevier—and by extension, academic journals—look less legitimate.

Predatory journals have the same effect: reducing confidence in the entire system. Even those who work in academia are sometimes unsure whether a particular journal is legitimate or not, and there are genuine borderline cases. For example, the *Frontiers* stable of journals appears (to me) to be legitimate, despite the fact that authors are required to pay a publication fee.[4] But some *Frontiers* journals appear questionable. *Frontiers in Public Health* controversially published articles linking vaccines and autism (Chawla 2016) and questioning the link between HIV and AIDS (Ferguson 2015); the first was subsequently retracted while the second was reclassified as "opinion." Perhaps in response to these incidents, the librarian Jeffrey Beall decided to add the publisher to his influential (but now sadly defunct) list of questionable journals (Bloudoff-Indelicato 2015). The controversy surrounding Beall's decision indicates how difficult such judgments are even for professionals. If it's hard for academics with expertise in the relevant fields to assess whether a particular journal or a particular publisher is legitimate, we can't reasonably expect ordinary people to make such judgments. If their confidence in scientific findings is lowered across the board as the result of such epistemic pollution, we can hardly blame them.

[4] Since conflicts of interest are a reason to discount expertise, it is incumbent on me to note that I have published in *Frontiers* journals on several occasions.

Epistemic pollution may be emitted by legitimate institutions of knowledge production, as well as from those who mimic such institutions. For example, it may arise from attempts to game systems that are supposed to track expertise. Institutions like universities, the bar association and peer review have as one of their functions the certification of expertise. But they have other functions, too, and these functions sometimes conflict. From this conflict, pressure to inflate credentials can arise. For example, universities have a financial incentive to overstate the expertise of their academic staff (thereby increasing their rankings, and attracting grant money and students). Systems that assess expertise may therefore be manipulated. There are many such cases: for instance, a Malaysian University was recently revealed to be urging faculty members to cite one another to boost citations (McCook 2017). Institutions, including the most prestigious, may be slow to investigate accusations of fraud or try to keep its discovery in-house, to protect the university's reputation.

To these sources of epistemic pollution we can add problems internal to the conduct of science, some of which have recently been widely publicized. Consider the so-called replication crisis, which we briefly discussed in the introduction. While much of the publicity to date has focused on social psychology, many of the problems seen in social psychology are just as common in other disciplines. For example, *publication bias* and the *file drawer effect* are certainly and notoriously problems in medicine. Publication bias is a kind of distortion in what gets published. It occurs when journals are more likely to publish certain kinds of finding than others, even though the intrinsic scientific merit of the favored kinds don't warrant the bias. Findings might be published because they are surprising, or because they're on certain topics (as Kitcher (1987) notes, it is far easier to publish work on human sexual behavior than on less exciting topics, and standards are accordingly lower). Perhaps the single biggest source of publication bias in science is a bias in favor of positive findings. Journals are full of papers that report that there is a significant correlation between two variables (framed as suggestive of a causal relation between them), or that a particular intervention significantly reduced the incidence of some pathology, and so on. Some of these findings are due to chance or would evaporate were

some factor controlled for; others are the result of what have come to be called questionable research practices (John et al. 2012) that torture the data for significance. These findings may go uncorrected due to publication bias: because papers that fail to replicate the finding are less likely to be published, or are published in less prominent places and fail to attract attention (see Fine 2013 for a discussion of how this dynamic plays out to amplify claims of sex differences and downplay evidence against such differences).

Publication bias may affect not only what gets published but also what research is conducted in the first place. Knowing that a failed replication will struggle to be published at all, and that if it is published the venue will not be high profile (and therefore will do relatively little to advance authors' careers) may discourage researchers from undertaking such research at all. For similar reasons, researchers may decide not to further pursue research when initial results are negative. This results in the file drawer effect: when negative results are filed away rather than submitted for publication. More pernicious still, researchers may repeat experimental protocols until they get the results they wanted. Selective publication of positive trials and suppression of negative findings may lead to an overestimation of the efficacy of new treatments (which may in fact be no better or even worse than currently accepted treatments). Unsurprisingly, this is a more common problem in industry-funded trials than in those conducted independently of industry (Every-Palmer & Howick 2014).

Industry funding is a general and central source of epistemic pollution. The tobacco companies, for instance, spent millions on funding research by university scientists. Surprisingly perhaps, the research they funded was often (though far from always) legitimate. The aim was often not to produce spurious findings but instead to draw attention away from tobacco as a cause of cancer by highlighting genetics, indoor pollution, and a host of other (in fact genuine but rare) causes. Industry funding also had an added benefit for the tobacco companies: the production and promotion of a cadre of (genuine) experts who were friendly to the industry and who could be called upon to testify in public forums in its support (Oreskes & Conway 2011).

Recent modeling work has demonstrated how powerful the promotion of genuine science can be in spreading epistemic pollution. Well-conducted science produces findings with a predictable statistical distribution: though good experiments are a reliable means to discover the truth, sometimes even the best experiments will throw up spurious results (that's why replication and supportive evidence from other kinds of scientific work are essential to the interpretation of science). Propagandists can take advantage of this fact to promote spurious findings.[5] Modeling work by O'Connor and Weatherall (2019) shows that even on the assumption that propagandists fund or selectively promote only genuinely well-conducted (but misleading) science, policy makers who attend to them can come to be more and more strongly persuaded of a false view, even as the scientific community converges on the truth; this remains true even if the policy makers also receive information directly from representatives of the scientific consensus.

Identifying Experts in a Polluted Environment

Goldman, Anderson, and other writers are optimistic that ordinary people can identify experts, using the criteria they set out. I think their optimism is misplaced. The epistemic pollution identified in the previous section makes the task of distinguishing reliable from unreliable sources too difficult for ordinary people to reasonably be expected to accomplish it.

The markers of expertise can play their certifying role only if they are not themselves excessively polluted. But these markers are polluted and they're known to be polluted. Ordinary people know that universities don't merely certify expertise. They know that universities also aim to attract funding and to manage public perceptions, and that these aims may conflict. Ordinary people know that peer review is conducted by people with their own interests and biases. They may reasonably (if

[5] Since small datasets are more likely to generate false positives and false negatives than large, the propagandists can ensure best bang for their buck by funding a larger number of smaller trials rather than the reverse (O'Connor & Weatherall 2019).

usually wrongly) conclude on that basis that certain views are not getting a fair hearing. Recall the example of the bacterial origin of stomach ulcers discussed earlier in this chapter. The medical community was slow to give the evidence due weight. Ordinary people may therefore wonder what other dissenting views don't pass peer review, due to the bias of reviewers, or what dissenting hypotheses are not investigated because granting agencies won't fund them. What should we make of the fact, for instance, that Michael Behe's own university department posted a disclaimer on their website, disavowing his "intelligent design" (a theory widely regarded as creationism in scientific dress)? Such a disavowal is predicted both by the view that Behe's claims are not well-supported by evidence, *and* by the view that scientists close ranks against dissenters.

Ordinary people assess expertise in an epistemically polluted environment, in which fakes, flakes, and frauds are promoted by merchants of doubt, by commercial interests and by a media in thrall to "balance" and to the need for sensation. They look to the markers of expertise to certify it, but these markers are themselves regularly manipulated. Moreover, they're aware that *all* sides—legitimate and illegitimate—are subject to extrinsic pressures. These worries affect every marker of expertise. Take track record. Cassam argues that David Irving's track record of misrepresentation ought to alert us to his deceptions. But track record is very often intrinsically difficult to assess. As we saw, Guerrero (2017) advises us to look to the *exoteric* record: predictions (or retrodictions) made by an expert that can be verified by the non-expert. Of course, sometimes experts (or putative experts) make predictions that can be easily verified or falsified. One well-known example concerns the neoconservative political pundit Charles Krauthammer. In response to the failure of the US military to find evidence of an active weapons of mass destruction program in Iraq—the ostensible existence of which had formed the central plank of his case for the war—he noted that the team had had only five weeks to find the WMDs. "Come back to me in five months. If we haven't found any, we will have a credibility problem," he wrote. He thereby provided an exoteric test for his credibility; a test he failed badly (Farrell 2013). But exoterically assessable predictions and retrodictions are the exception, not the rule. In many areas, what exactly

is being predicted and how it would be falsified is often hard for the non-expert to assess.

In specialist science, it often takes specialist knowledge to understand just what is being predicted, let alone to verify the prediction. Climate denialists, for example, seem to be committed to the prediction that global temperatures will not correlate with concentrations of CO_2. Climate scientists will tell you that this prediction has been falsified: as a matter of fact, temperature rises are well correlated with CO_2. But the denialists have a response, or a number of responses. They may maintain that the apparent correlation reflects manipulation of the data, rather than genuine change. It has been alleged, for example, that the National Oceanic and Atmospheric Association tampered with temperature data inconvenient to the warming narrative (Richardson 2017); that the Climate Research Unit at the University of East Anglia deleted its data to hide anomalies, and so on. All of these claims have been multiply debunked, of course. The interested reader would do well to consult the blog *Skeptical Science* for both entry-level and advanced discussions of all these myths.[6] But these debunking efforts have themselves been met with (attempted) debunking.

Liberals tend to think of denialists as ignorant, unintelligent or hucksters. In fact, sophisticated denialism is easy to find. *Watts Up With That*? describes itself as the "The world's most viewed site on global warming and climate change" and promotes denialism with (apparent) facts. Judith Curry, a climate scientist with a solid track record of well-regarded publications, uses her scientific skills to promote doubt on her own blog (as well as the blogs of others, in the media, and in front of US House committees). Assessing her claims, as well as those of other sophisticated denialists, is far beyond my capacities. Are the empirical claims she makes true (e.g., that a particular technique has been misapplied, or that there are statistical errors in a paper)? Assuming they are true, do they support to some degree her skepticism regarding the scientific consensus? While I find the responses on *Skeptical Science* more plausible than her posts, it's likely that my disposition to defer to those on my side helps explain that fact. I'm confident this point generalizes:

[6] https://skepticalscience.com/

our being swayed by the arguments is due in part to our responding to other cues. I'm not of course claiming that there's no fact of the matter here, or that one side doesn't have very much better evidence than the other. I'm suggesting, rather, that *we* (non-specialists) lack the capacity to identify which side is right by reference to argument quality alone (or, typically, even primarily).

Argumentative capacity—the possession of what Goldman (2001) calls "dialectical superiority"—fares no better as a marker of expertise (D. Coady & Corry 2013). The ability to rebut arguments and the *appearance* of having this ability may dissociate. As many scientists who have debated creationists have learned to their cost, well-rehearsed debaters can seem to neutral audiences to be dialectically superior by having an apparent response to every objection, even if the response is only smoke and mirrors. They can also appear to evince dialectical superiority by raising so many objections and making so many points so quickly that their opponent is unable to rebut more than a small fraction of them (this is known as the Gish gallop, after a creationist who specialized in the technique). Reference to dialectical superiority enables us to distinguish those who have spent a lot of time on a topic from those who haven't, but it's insufficient to allow us to distinguish genuine experts from pseudo-experts who have also spent a great deal of time on the topic.

Intellectual honesty fares no better, because it's not appropriately independent of the issues disputed by the experts themselves. All sides accept, with Anderson (2011), that a putative expert acts dishonestly if she doesn't withdraw claims that have been refuted. But the fakes and fellow-travelers hold (sincerely or duplicitously) that it's the genuine experts who are intellectually dishonest because it is *their* claims that have been refuted. Similarly, accusations of conflict of interest are often unhelpful, because (as Guerrero notes) such conflicts typically appear on all sides. It is of course common for anti-vaxxers to accuse their opponents of being in the pockets of "big pharma," and climate denialists cite the attractions of grant money to explain the appearance of consensus among scientists. Both can point to genuine scandals in the relations between scientists and pharmaceutical companies, such as the phenomenon of medical ghost-writing (Langdon-Neuner 2008), where

a prominent physician or researcher puts their name to a paper that has been largely or even entirely written by company representatives.[7]

Claims concerning the existence of a consensus on a topic are also of limited help, insofar as such claims also fail to be appropriately independent of other issues. If credentialing bodies will not grant PhDs to dissenting researchers, for instance, we should expect to see a consensus of appropriately credentialed scientists on a topic, regardless of whether the consensus is well supported. If data that conflicts with the consensus view is suppressed (deliberately or just because it's difficult to publish), the consensus will not have much evidential value. This point is a generalization of Goldman's claim that the concurrence of additional experts with a claim adds no additional evidential weight to it unless they are sufficiently discriminating in what they believe. Goldman worries about excessive deference to opinion makers, but there are other ways in which an unreliable consensus could be generated. If institutions that grant credentials use inappropriate criteria in assessing expertise, the resulting consensus will not be truth conducive.

Claims of intellectual dishonesty are also symmetrical. Climate scientists routinely (and to my mind rightly) accuse some of the denialists of deliberate deception.[8] Sometimes, persuasive evidence of such deception emerges, when memos and emails never intended for public consumption comes to light (such as the tobacco industry memo that provided *Merchants of Doubt* with its name). But such revelations are rare, and *their* side thinks that they have such smoking guns too. The most famous here is "climategate." In November 2009, a server at the Climactic Research Unit at the University of East Anglia, was hacked

[7] One of the clearest cases of intellectual dishonesty in recent medical history is surely the Andrew Wakefield story. In 1998, Wakefield and his co-authors published a paper alleging a link between the MMR vaccine and autism. After other researchers failed to replicate his findings, Wakefield was found to have undisclosed conflicts of interest. The British General Medical Council then investigated further and found a litany of other problems, from performing unnecessary and invasive procedures on children with autism to suppressing data. The paper was retracted and Wakefield was struck off the medical register. Those who trust the relevant institutions will take Wakefield to be discredited and his research invalidated. But if you are disposed to distrust these institutions, you might see them as closing ranks against a brave truth-teller.

[8] Only some. Some of the dissenters are sincere, and some of the sincere dissenters are competent. Indeed, their dissent may be explained, in part, *by* their competence: their mistake (or one of them) is to overestimate the powers of individual human reason.

and a trove of documents and emails stolen. Excerpts from the emails were subsequently published on denialist blogs, which alleged that they show that researchers were fabricating and manipulating data to support their political line. Denialists seized on one email in particular, in which Phil Jones (a leading climate scientist) said he used "Mike's *Nature* trick…to hide the decline." Here was the smoking gun! The mainstream media reported the revelations with varying degrees of credulity: for the *Telegraph*, for instance, it was "the worst scientific scandal of our generation" (Booker 2009). Three separate inquiries found no evidence of scientific misconduct. Unsurprisingly, the denialists allege the inquiries were themselves fraudulent. What would you expect when the scientists close ranks?

For each claim by a scientist or a group of scientists, it seems that there is a rebuttal by an opponent, a response to that rebuttal and a further response to it in turn. Good luck keeping up! While one side may feature better credentialed experts than the other, such a pattern of distribution is exactly what one would expect if the better credentialed side suppressed dissenting research (such suppression might be conspiratorial, but it need not: it could even be produced by well-meaning but biased scientists trying and failing to give their opponents a fair hearing). Claims of intellectual dishonesty abound on both sides, but for the most part they don't help, because the accusations are symmetrical and we can often adjudicate the claims only by adjudicating the first-order issues on which they turn. We can't, therefore, utilize concerns about intellectual dishonesty to identify reliable experts: the criteria aren't sufficiently independent of one another.

The Efficacy of Epistemic Pollution

So far I've been long on argument and assertion about the effects of epistemic pollution, and short on empirical evidence. We are, as I've already noted, epistemic individualists, and we tend to be confident of our intellectual powers. Readers of a book like this one are particularly likely to have a high (and probably well-founded) opinion of their capacities.

Surely I exaggerate the degree to which epistemic pollution is an obstacle to belief? Surely you (dear reader) can, with sufficient effort and application, sort through the lies and the fog, and come to an accurate assessment of the evidence?

You are (very probably) in a much better epistemic position than most people. It's not just that you are well-educated and (again, very probably) more intelligent than average. It's not just that you probably have research skills that most people lack. You are also (very probably) epistemically luckier than most. As a consequence of your socialization (from family through to prestigious academic institution), you have acquired dispositions to trust reliable sources. You know enough to distinguish legitimate institutions from diploma mills; you have some idea of the degree of legitimacy conferred by a publication in *Nature* or *Science*. You are alert to signs of predatory publishers and on the lookout for industry funding. You are therefore protected, to some degree, from epistemic pollution.

For all these reasons, you're indeed more likely than most to get things right when you (attempt to) judge for yourself. But that's not because you're a counterexample to my claims: it's because you fit my model so well. It's because you defer well that you do well. When you attempt to judge for yourself, you actually engage in social cognition; and that's why you tend to get things right. You can reliably adjudicate between David Irving and his many critics, between climate scientists and denialists, between anti-vaxxers and genuine experts. But while it may seem to you that you do so well (epistemic individualist that you are) through the power of your unaided reason, a very important part of the explanation for your success is that you defer so fluently and appropriately. You owe your success to the way in which you are embedded in epistemic networks.

Even so, I bet even you sometimes go wrong. Your capacities, and your disposition to defer, only get you so far. You live in an environment that is unreliable, in which frauds and fakes mimic the cues to reliability you rely on. Sometimes—I bet—you fall for their tricks. I certainly have.

One of the examples featured in *Merchants of Doubt* is "acid rain." The phenomenon was first recognized in the mid-nineteenth century,

and regulations were introduced by the British parliament to address it. The problem returned to scientific and public consciousness in 1974, when Likens and Bormann (1974) published a paper in *Science*, showing that acid rain was a serious problem in large areas of the United States. The National Academy of Sciences and the EPA both launched investigations, and both concurred: acid rain was a "serious hazard to human health" (O'Connor and Weatherall 2019, 37). The Carter administration moved to regulate the power plant emissions that were largely responsible. But implementation was left to the incoming Reagan administration. And then the merchants of doubt moved in.

There's no need to rehearse, here, how some of the very same people who had been involved in defending tobacco and who would later obfuscate the science of climate change hijacked the process and ensured that it stalled. *Merchants of Doubt* tells the story much better than I can. Here I want to mention just one thread in the broader narrative. Edward Krug, a soil scientist at the Connecticut Agricultural Research Station, was promoted by the denialists and by institutions sympathetic to them as offering a view contrary to those who called for increased regulation. Krug argued that changes in soil acidity were largely the product of natural processes, not acid rain (Krug & Frink 1983). His claims were assessed and quickly refuted: acid rain was swamping such processes (Galloway et al. 1984). But despite the refutation, Krug's work was picked up beyond the science journals and weaponized in the fight against regulation. It was presented in *Policy Review*, *Reason Magazine*, and even on *60 Minutes*. The pollution seeped in everywhere: in 1990, NPR reported that the scientific consensus on acid rain was that the issue was complicated.

Surely, though, sophisticated people, people with a background in science and who are responsive to the right cues and read the right sources, surely *they* are able to see through the fog? In *Merchants of Doubt*, Naomi Oreskes has a confession to make: in the early 1990s she "used Krug's arguments in an introductory earth science class at Dartmouth College to teach 'both sides' of the acid rain 'debate'" (Oreskes and Conway 2011, 103). In an epistemically polluted environment, even the most sophisticated people risk being taken in.

Restoring Trust in Science

If we're to bring people to believe better, it won't be by asking them to behave more responsibly or by inculcating the epistemic virtues in them; not primarily and—I bet—not very importantly either. Epistemic humility, open-mindedness, care in evidence-gathering—these all good things (in their place). But they're no solution to the problem of believing better, largely because it's extremely difficult, and perhaps impossible, reliably to judge when they're called for and when they're not. They're dispositions that can as easily lead away from the truth as toward it (Levy & Alfano 2019).[9] More pointedly, it's simply false that the epistemic virtues and their responsible application enable the person reliably to track truths. To the extent she succeeds, it is her embedding in appropriate epistemic and social networks that enables her success.[10]

[9] Those people who generate conspiracy theories—as opposed to those who consume them—exhibit a great many of the epistemic virtues (K. Harris 2018). Conspiracy theories often begin from the identification of an anomaly: a piece of data that appears to conflict with the official explanation. The theorist exhibits open-mindedness in looking for an alternative explanation. She looks to alternative epistemic communities, thereby displaying epistemic humility. She certainly can't be faulted for a failure to look for evidence: conspiracy theorists may be voracious in their consumption of reports. In many ways, conspiratorial ideation looks like science: it is science gone wild, science no longer constrained by the epistemic networks within which mainstream scientists work.

[10] Eric Winsberg (2018), who accepts (and strengthens) the case for the claim that nonexperts can't hope responsibly to assess climate science for themselves, nevertheless argues that individuals can and should engage in careful assessment prior to accepting consensual scientific claims. We should assess not the basis of the claims the science makes, but the structures that underlie the generation of the consensus. Indeed, Winsberg thinks it's uncommonly *easy* to engage in this kind of assessment in the case of climate science. This is for several reasons. First, climate science is the product of a multiplicity of different disciplines and this fact ensures robustness against corruption in one area of the science and cross-checking of findings. Were one discipline's contribution suspect, the others would detect the problem when its results impinged on theirs. Second, climate science has an institution—the IPCC—that summarizes and assesses the science. Third, climate science is subject to well-funded hostile scrutiny. These facts ensure that the consensus is robust, and these facts are easily discerned by laypeople.

While Winsberg is surely right that these facts about the structure and the institutional setting of climate science entail a high degree of credibility, I am both much less skeptical that consensus (however generated) is good evidence and much more skeptical of ordinary people's ability to discern and understand the facts about structure that help to ensure reliability. Climate science *denial* is also the product of multiple disciplines. It, too, has institutions (including a shadow IPCC) that claim to summarize its results and identify the work that is reliable. It, too, is subject to well-funded hostile scrutiny. Just as markers of expertise can be mimicked, so can institutions and structures. Of course, the merchants of doubt haven't been able to mimic the extent and depth of the structural and institutional network underlying climate science. But I only know *that* by testimony! I don't know how to verify these claims in any

Philosophers often advocate the teaching of critical thinking skills, such as the capacity to identify argumentative fallacies, as a partial solution to our epistemic crisis. But such teaching has small and short-lived benefits (Mercier et al. 2017). A broader general education, including scientific education, also doesn't seem to reap any benefits. In fact, it may hurt: as we've already seen, better educated Republicans are *less* likely to accept the consensus view on climate change than less well educated Republicans (Kahan 2015). Better educated Republicans are also more likely to think that Obama is a secret Muslim (Lewandowsky et al. 2012). Better education and more tools for argumentation may enable those who distrust the institutions of science and the universities to counter their claims more effectively. This may arise from what Taber and Lodge (2006) call the sophistication effect, whereby being more knowledgeable provides more ammunition (and more skills) with which to counter unpalatable claims.

The restoration of trust in science and scientific institutions is likely to make a bigger and longer lasting difference to the goal of better belief formation. On the right, trust in these institutions has ebbed significantly in recent decades (Gauchat 2012). This distrust has generalized to the universities as a whole: a majority (58 per cent) of Republicans now say that colleges and universities have an overall negative effect on the United States (compared to 19 per cent of Democrats). The same survey shows that 85 per cent of Republicans have a negative view of the news media (Doherty et al. 2017). If we are to promote better belief, we need to promote better deference, and that requires the restoration of trust in these institutions. Central to doing so is reduction of epistemic pollution.

Epistemic pollution *rationally* reduces trust in institutions. If you know that the same institutions that credential science—universities, for

other way and I'm skeptical there *is* any other way. If testimony is good enough for knowledge about the structure and institutional setting of climate science, then it's also good enough for acquiring knowledge about the claims of climate science.

I'm also skeptical that knowing the facts about the institutions of climate science positions me to assess the degree to which these institutions are knowledge-conducive in any case. What counts as a knowledge-conducive structure and what counts as group think or corruption is a difficult issue, and not one that most people can assess for themselves. I am, in fact, skeptical that *anyone* can accomplish this on their own.

example—are also involved in gaming the credentialing system, your trust in them is rationally lower than otherwise. If you know that the funders of science often have conflicts of interest that might (and do) lead them to suppress unfavorable data, then you should reduce your trust in them. If you know that scientists themselves sometimes engage in questionable research practices, you should be somewhat slower to endorse their findings. While we can't realistically hope to eliminate these practices, we can and should take steps toward their reduction.

Pollution of the traditional kind is often hard to tackle due to a collective action problem: while everyone might be better off if no one pollutes, no individual can make a significant difference on their own, and any individual who pays the cost of clean-up locally is worse off than others who don't cooperate. Collective action problems are solved by mechanisms that ensure that (almost) everyone contributes to the goal. There are multiple ways this can be done, but often (and especially in cases when some of the actors don't share the goal), some degree of coercion is required. Epistemic pollution is also a collective action problem: while most of us would be better off if it were significantly reduced, individuals can't make a significant difference to it by themselves, and anyone who acts alone is worse off than others who don't cooperate.[11] It is, moreover, a collective action problem made worse by the fact that some actors don't share the goal most of us would like to achieve: merchants of doubt, purveyors of predatory journals and peddlers of expensive and ineffective drugs may prefer to go on polluting to having a clean epistemic environment. Reducing epistemic pollution will almost certainly require some degree of coercion, from government or other institutions with the clout to impose costs on those who don't cooperate.

While I'm not the right person to develop policy proposals, some preliminary steps toward restoring trust are obvious. We need to vastly reduce the number of predatory or fake journals, or ensure that such journals are effectively confined so that they aren't (and aren't seen to be) contaminants in the scientific ecosphere. Doing this requires that legitimate open access journals are clearly distinguishable from

[11] Everett and Earp (2015) suggest that the replication crisis is a tragedy of the commons; I think it's plausible to generalize this point across a range of epistemic pollutants.

illegitimate. This is a task for the scientific community as a whole. Universities should refuse to pay publication fees for journals identified as illegitimate and researchers who publish in them should not receive credit (in the form of citations, promotions or grant funds) for such publications. Such a move would starve the illegitimate journals of funds and should lead to the closure of most. Beall's list was a great start, but it would be better done collectively.[12] As we saw above, the list was controversial, in part because it erred on the side of considering journals illegitimate. It would be better to use less-demanding criteria. A consensus of the scientific community on the vast majority of such journals could easily be reached and a very significant epistemic contaminant drastically reduced.

Problems in the conduct of legitimate research must also be addressed. Incentives should be put in place for the replication of research; such incentives should be combined with a greater willingness to publish failed replications (alternatively, institutions can mandate such replications as part of the training of graduate students; see Everett and Earp (2015) for a proposal along these lines). Hypotheses and methods should be preregistered to ensure that researchers don't engage in questionable practices *post hoc* to ensure significance. Preregistration also eliminates the temptation for selective reporting of results: if only and all preregistered studies are published, we can be confident that we have the full array of data. Statistical techniques can be utilized to compensate for the file drawer effect and thereby generate more realistic effect sizes. Such techniques can also identify evidence of data manipulation, such as p-hacking. These proposals are by no means novel: in fact, many are already being implemented. Prestigious journals in psychology, for example, have implemented changes to their practices, requiring bigger sample sizes (lowering the risks of chance findings) and

[12] Beall's list of Predatory Journals and Publishers was maintained and updated by Jeffrey Beall, an academic librarian, from 2008 to 2016. Though controversial, it was widely respected and consulted. While it remains unclear why Beall chose to shutter the service, there is evidence that pressure from predatory publishers played a part in his decision (Straumsheim 2017). Individuals are less able to resist such pressures than collectives, in which pressures can be shared and mutual support offered. Moreover, collective decisions may be less controversial, especially if the decision-making body includes individuals with different perspectives and interests.

encouraging preregistration of hypotheses and methods (see Lindsay 2015). If the more prestigious journals all follow suit, ambitious scientists will be forced to adopt these standards and the degree to which science is unreliable should fall.

We should also reduce the incentives for science by press release and the extent to which new research is presented in the mass media (and to a lesser extent in the journals themselves) as revolutionary and earthshattering. This, too, is a collective action problem. Researchers likely prefer a world in which everyone refrains from hyping their research to the current situation.[13] Most might also prefer that media attention was not a significant determinant of prestige, promotions, and grant success. But given that media attention is valued by institutions and granting agencies, and no individual researcher can change the culture by themselves, each feels that they have to play the media game, which means representing their research as more important and revolutionary than it really is. The result is that consumers of the media are left with the impression that yesterday's research findings have been overturned by today's, and that today's will be overturned by tomorrow's—all of which has the result of reducing trust in any particular finding or claim.

Of course, actually implementing the agreements needed to solve collective action problems is difficult, especially given that science is an international enterprise. There are at least two possible routes to effective regulation. One is through governmental action: if the United States and the EU ensured that funding of science was tied to responsible media engagement, norms might change across science (given the proportion of science funded by them). Bringing China on board would be even better, and might be possible. Admittedly, in many domains the record of government is not encouraging: when questions come to be politicized, policy often ignores expert opinion. We can reasonably be more optimistic about self-regulation from within epistemic

[13] What if the research genuinely is earthshattering? I strongly suspect that rules regulating science and its reporting should not be written in ways that make explicit allowance for such eventualities. The genuinely earthshattering is sufficiently rare that we do better to design regulations that assume that the research governed by them is normal, not revolutionary, science. If research is sufficiently significant, this can be expected to be evident without the need to hype it.

communities, insofar as policy makers within these bodies must remain responsive to the expert opinion of their members. National peak organizations, for instance, could regulate the conduct of science in ways that could produce the same effects (self-regulation by epistemic communities may also reduce worries about overreach, given that such communities have a much smaller domain over which they exercise power than do governments). Perhaps no official mandates are necessary: norms within science are changing rapidly, and many researchers now look askance on unregistered hypotheses and small numbers of participants. Social pressures may go a long way toward fixing many problems.

Changing media norms is a tougher nut to crack. In the contemporary environment, the media is fractured and cooperation unlikely. The collapse of traditional funding models has left media organizations—where they survive at all—chasing clicks, which encourages sensationalism. It might be possible to produce many of the desired effects without the media on board: if reputable scientists withdraw cooperation with sensationalistic media, they may come to be known to feature only charlatans and the likelihood that the public will ignore them will rise.

None of the measures mentioned above, some of which have already been introduced (albeit patchily), would solve the problem of distrust in science in the short term. When trust is lost, it's difficult to restore, and measures taken by the very institutions distrusted are likely to be regarded with a skeptical eye. Over the longer term, however, removing epistemic pollutants from the environment should increase trust in reliable sources of information, and thereby improve belief formation.

In this and the previous chapter, I've argued that individual cognition—unaided—is much less powerful than we tend to think. Without heavy duty scaffolding and heavy reliance on others, we're very much less reliable than we hoped, even in the best cases. Worse, we're rarely in the best cases: we live in an epistemically polluted environment, in which others seek to misdirect us. When we do well (as we frequently do), we tend to attribute our success to our own individual cognition, but that's only because our deference is so smooth and automatic, we fail to notice it.

Showing that thinking for ourselves is less powerful than we thought or hoped is one thing, however; it's quite another to show there's an

alternative that is both more successful and also ethically permissible. It's to that task that I turn next. The epistemic engineering I advocate is, or is closely akin to, nudging, and nudging is hugely controversial. It's widely seen as impermissible, or at least undesirable, on the grounds that it subverts individual autonomy. I'll argue that while nudging *can* be used to subvert autonomy, that's not because there's any distinctive problem with influencing people's behavior in this kind of way. Nudging is the presentation of information, and done appropriately it's no more subversive of autonomy than is giving (explicit) arguments for a conclusion or a course of action. In making the case for the permissibility of nudging, I'll also complete my case for seeing ourselves as rational animals. I'll provide a fuller defense of a claim I've made several times already: that deference is fully rational: not merely *ecologically* but also *directly* rational.

6
Nudging Well

Roughly, a nudge is a way of influencing people to choose that works by changing aspects of the "choice architecture" (Thaler and Sunstein 2008, 6)—the context in which agents choose—to encourage better choices (better usually insofar as these choices promote the welfare of the choosing agent herself; occasionally nudges aim at the promotion of social welfare instead of individual welfare). It's a familiar fact that people often make choices they themselves recognize are in not in their own interests. Nudging can bring them—us—to choose better.

For example, people often have unhealthy diets. Making healthier foods more cognitively accessible (for example, by putting them at eye level) increases their consumption relative to less healthy foods (Rozin et al. 2011; see Bucher et al. 2016 for review). Similarly, people often fail to save an adequate amount of money to fund a decent retirement. They can be nudged into saving more: if the default option presented to new employees sequesters a higher percentage of salary to a retirement savings account, people tend to save more, because they tend to accept the default (see Smith et al. 2013 for review). Options can be *framed* in ways that change behavior: agents are risk averse when options are framed positively, but risk seeking when options are framed negatively, even though the options may be identical in arithmetical terms (Tversky & Kahneman 1981).

While nudges may be in the interests of individuals, however, they're extremely controversial. Critics often argue that nudging *manipulates* us (Bovens 2008; Wilkinson 2013; Saghai 2013). Nudging threatens our autonomy because autonomous choice is rational choice, and nudging bypasses our capacities to reason.

In this chapter, I'll argue that nudging doesn't manipulate us. Nor do nudges bypass reasoning. Instead, nudges work by providing genuine evidence to agents, and when they change behavior, the change occurs

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0006

in response to this evidence. Insofar as I advocate improving belief formation by thoroughgoing engineering of the epistemic environment—not merely by clamping down on pollutants, but also by altering cues to belief—the success of this argument is essential for my project. At very least, it would be a large cost to its acceptability if I were forced to advocate manipulation on a large scale. But I have an even broader aim in mind. Coming to see how nudges work allows us to better understand the mechanisms that underlie the processes discussed in previous chapters (social referencing and reliance on environmental cues, for example) as themselves reasons-providing. They're ways of changing minds through the provision of higher-order evidence.

Nudging and Autonomy

Autonomy—the capacity of an agent to govern herself—is highly prized by us. Autonomy is nowhere more significant than in the life of the mind. Kant's injunction *Sapere aude!* (have the courage to use your own understanding) is motivated in important part by the connection between using one's own understanding and autonomy. Of course, a major theme of this book is that individual cognition is much less powerful than we think, and that our epistemic capacities are very importantly owed to the distribution of cognitive labor instead. In light of this fact, and the link between the use of one's own understanding and autonomy, it might be thought that I would respond to the worry that nudges threaten our autonomy by arguing that autonomy isn't so valuable after all; that we overvalue it because we overvalue individual cognition.

This isn't in fact a thesis I want to defend. While it may be true that we overvalue autonomy, due to our overvaluation of individual cognition, autonomy seems to me to be genuinely worth defending. There are at least two reasons why autonomy is genuinely valuable. First, agents are often in a better position than others to make decisions concerning the shape of their own lives. It is one thing to think (as I have argued we should) that we ought to defer to epistemic authorities about difficult and complex questions and quite another to think that we can't come to

reliable beliefs about ourselves. Perhaps even in the personal sphere the social sciences have much to teach us, but we retain *some* degree of epistemic privilege with regard to our own values and preferences.

Second, even if we're apt to decide some questions badly, we seem to have a *right* to settle the shape of our own lives and to decide on a conception of the good for ourselves. When the harms we foreseeably risk will fall primarily on ourselves, others have little or no right to override our decisions. We want our lives to be our own, reflecting our own values and priorities, and even our own mistakes. The value of autonomy may often be overstated but I'm not ready to abandon it. Any impingements on our autonomy require justification.

But nudging *does* seem to threaten our autonomy. In fact, both advocates and opponents of nudges accept that it does. Nudging seems *paternalistic* (Thaler and Sunstein, the original advocates of nudging, describe their program as 'libertarian paternalism'): it manipulates us into making decisions in our own best interests, rather than leaving us to make decisions for ourselves on the basis of our own reasons. There's nothing wrong with parents making decisions on behalf of their children, when (and because) the children lack the rational capacity to make these decisions for themselves. But fully rational agents rightly value making decisions for themselves, and nudging bypasses our capacities for rational agency.

To see how nudges are supposed to bypass rational agency, let's look at a possible nudge in action. An unscrupulous (or perhaps a well-meaning) election official might nudge voters in an upcoming election to vote for a particular candidate by listing that candidate first on the ballot. The official would be taking advantage of the *ballot order effect*: the small but sometimes significant advantage that accrues to names listed higher on the ballot paper (Darcy & McAllister 1990; King & Leigh 2009). The ballot order effect serves as John Doris' prime example of how these (supposedly) non-rational influences may threaten autonomous agency (Doris 2015, 2018). As Doris emphasizes, that a candidate is positioned higher on the ballot is not a genuine *reason* to favor that candidate. Ballot order doesn't correlate with candidate quality (in most jurisdictions, name order is determined by drawing lots). It follows that in being influenced by ballot order, a person has their choices

shaped by facts that are *not* good reasons. To the extent that choice is influenced by ballot order, genuine reasoning is bypassed. Doris therefore insists that these kinds of influences don't involve genuine reasoning or genuine reasons. These influences are, he says, "deeply unintelligent."[1]

I treat you as a rational agent and respect your intellectual autonomy when I give you *reasons* why you should vote for one candidate rather than another. If I nudge you into voting for someone by listing them first on the ballot (or, say, making a vote for them the default option) I do neither. I bypass your capacity to deliberate and instead appeal to "deeply unintelligent" mechanisms. It's because nudges appear to influence choice without offering reasons that both opponents and proponents of nudges believe that nudging is paternalistic. Opponents point to this (alleged) fact as a major plank in their case against nudging, while proponents cite the inevitability of nudging in its defense. Thaler and Sunstein accept that nudges take advantage of the fact that we are—in their own words—"somewhat mindless, passive decision makers" (2008: 37), but argue that there's simply no alternative to nudging. The deeply unintelligent mechanisms that respond to nudges are ubiquitous and nudging is inevitable. If we don't nudge deliberately, people will be nudged nevertheless, either by bad actors who seek to manipulate them or by chance. Whatever we do, or fail to do, we'll all be nudged nevertheless; we might as well put nudging to good use.

Whether nudging is really inevitable and whether there's a normative difference between being intentionally and unintentionally nudged (as Alfano (2013) and Kumar (2016) each argue) are interesting questions, but they're not questions we need answer here. Nudges don't simply manipulate us by bypassing our capacities to reason. Instead, they provide us with evidence, which we typically weigh appropriately. Nudges don't tend to provide arguments or evidence that fit our paradigms, but that's because our paradigms are of *first-order* evidence. We neglect *higher-order* evidence, but higher-order evidence is genuine evidence.

[1] Doris is here quoting Stanovich (2004). But Stanovich is more careful than Doris (unsurprisingly, given he has long resisted the facile identification of type 1 processes with irrationality (Stanovich 2018)). Stanovich writes that these processes are "*in some sense* deeply unintelligent" (39; emphasis added); Doris drops the qualification.

Nudging Higher-Order Evidence

First-order evidence, our paradigm of evidence, is evidence that bears directly on the truth or falsity of a particular proposition. The pattern of blood spatter in the room is evidence that the killer used a knife; the fingerprints on the light switch are evidence that the killer was the butler.[2] Higher-order evidence is evidence about our evidence. In epistemology, the main focus of debates about higher-order evidence has been the reliability of the agents who assess the evidence; in particular, on how disagreement can provide evidence about such reliability. Consider this (by now hackneyed) case, *Restaurant Check*:

> Suppose that five of us go out to dinner. It's time to pay the check, so the question we're interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20 percent tip, and we further agree to split the whole cost evenly, not worrying over who asked for imported water, or skipped desert, or drank more of the wine. I do the math in my head and become highly confident that our shares are $43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are $45 each
>
> (Christensen, 2007: 193).

Most philosophers agree that under certain conditions, a disagreement of this kind provides higher-order evidence, and that evidence puts rational pressure on the parties to reduce their confidence in their calculation. If the agents who disagree are *epistemic peers*, then they should each conciliate (i.e., lower their confidence in their judgment). Peer dissent is evidence for each person that at least one of them has made a mistake; given that neither has a reason to think that the other is more likely to be in error than themselves, they should treat the disagreement as evidence against their conclusion. It's not first-order evidence (not evidence that they failed to carry the 2 or that they left off one item). It's

[2] Most epistemologists prefer to talk about our mental states, rather than the objects of these states, as our evidence. As far as I can see, nothing of significance turns on which approach we take in this context.

higher-order evidence: evidence that they may not have processed their first-order evidence well.

Given that disagreement is ubiquitous, if we must conciliate whenever we encounter it, we seem to lose the right to confidence very broadly. Conciliationism thus gives rise to what has been called the problem of *spinelessness* (e.g., Elga, 2007; Fritz, 2018). Many epistemologists respond to this worry by defending an extremely demanding account of epistemic peerhood, according to which my peers have exactly the same evidence and the same capacities as I do.[3] Since each of us has few peers (so defined), we are each under less pressure to conciliate. As Jennifer Lackey (2010) points out, this maneuver risks cutting the debate over the epistemic significance of disagreement off from the real world cases of dissent that motivated interest in it in the first place. In fact, higher-order evidence, of various strengths and kinds is ubiquitous and we are pervasively and appropriately responsive to it.

In real-life analogues of cases like *Restaurant check*, we rightly treat disagreement as high-order evidence without needing to know the track record of a dissenter. The fact that a sober, apparently well-functioning adult disagrees with me about a sum that's difficult enough for a mistake to be unsurprising is *some* evidence that I have made a mistake. Even if I know that the agent is *not* my peer, in the exacting sense common in the literature, his dissent is still higher-order evidence against my conclusion: I can't entirely dismiss his dissent on the grounds that I'm (say) 5 percent more likely to be right than he is. Of course I can dismiss his dissent if the sum is trivially easy (Lackey discusses a case in which a dissenter disagrees on the sum of 2+2) or if he lacks the competence to perform a calculation like this one. But dissent quite routinely provides some degree of higher-order evidence.

[3] Setting the bar for peerhood extremely high is not the only response to the problem of spinelessness. One influential response turns on the attractiveness of a test for peerhood along the following lines: my peers are those agents who are as likely as I am to be right about the issue under dispute. Combined with the so-called *independence principle*, according to which an agent's reasons for discounting a dissenting peer's opinions must be independent of the dispute itself, the test allows may allow us to hold fast to our controversial opinions. *Setting aside the dispute and all the reasons implicated in it*, we have no basis for thinking that dissenters would be as likely as we are to come to the right response (Elga 2007; Fritz 2018; McGrath 2008).

For that matter, *agreement* also provides higher-order evidence. Given that a calculation is moderately difficult for me, if I come to the same answer as an independent agent I should raise my confidence in it. The greater the likelihood that I might have been mistaken, the stronger the evidence provided by independent agreement. The *number* of others who have independently tackled the problem should also make a difference to my confidence. If I'm the lone dissenter at a table of 8, my confidence in my answer should be low; conversely, if many others agree with me (and few disagree), my confidence should rise. Numbers make a difference for two reasons. First, the likelihood that I've made a mistake rises or falls as a function of the number of others who agree with me: the higher the proportion of agents who agree with me, the lower the likelihood that I've made a mistake. Second, sheer numbers make a difference to the plausibility of an appeal to what Lackey (2010) calls "personal information," such as my knowledge that I'm paying attention and I'm being sincere, to break the symmetry between me and dissenters. Again, the higher the proportion of dissenters, the more implausible an appeal to such information to dismiss them, at least when the dissenters are to some significant degree independent of one another.

Of course, all this is true only if other things are equal. Experts ought to give little or no weight to dissent when it comes from those who lack expertise. An expert on climate change shouldn't lower her confidence in her predictions and her models because Donald Trump declares that global warming is a hoax. Nor should she be impressed by the enormous number of dissenters, given that she's an expert and almost none of them have any of the specialist skills to understand her work.

With these facts in mind, we can begin to glimpse the ubiquity of higher-order evidence. We can also begin to see how often I've appealed to it throughout this book. In the previous chapters, for instance, I discussed the role that markers of expertise (possession of relevant qualifications, of a track record of publication, of prizes and citations, and so on) should and do play in guiding our response to testimony. In appealing to such markers, I appealed to higher-order evidence. In giving more weight to some opinions on the basis that they come from someone who possesses these markers, we are taking higher-order evidence in favor of

their views into account. In appealing to an expert consensus, we're also appealing to higher-order evidence.

Markers of expertise are just one of the more obvious kinds of higher-order evidence I've appealed to in this book. Once we've seen how nudges, too, provide higher-order evidence, we'll be in a better position to see just pervasive such appeals have been—that is, we'll be in a position to begin to glimpse just how much we lean on higher-order evidence in ordinary and expert cognition, and how important such evidence is in our epistemic lives.

Nudges as Evidence

How do nudges work? Exactly how they cause behavior, on the standard understanding of their influence, remains elusive: theoretical models often invoke vague notions like "salience," which seem more like placeholders for mechanisms than explanations. There is, however, more or less universal agreement that however they work, they bypass rational cognition. While I don't claim to be able to do very much better at providing a proper account of how they function, I suggest that nudges *do not* bypass rational cognition. Instead, at least typically their influence is due to the manner in which they provide implicit recommendations, and therefore higher-order evidence in favor of the option nudged.

Let's begin with the ballot order effect, which was Doris' prime example of how these kinds of influences may threaten autonomy. As we saw, the ballot order effect is supposed to be irrational, because candidate order doesn't correlate with the quality of candidates. Of course that's true: since candidate order is settled by lot or in some other way that doesn't track quality (e.g., by the order in which they registered for the election), order doesn't provide *reliable* evidence in favor of any candidates. But it doesn't follow that it doesn't provide evidence at all. The order in which items are listed implicates their importance (Green 1998). Think of how news is presented online or on TV. The most important information is presented first ("our headlines at the top of the hour…"); being listed first is therefore implicit testimony that an item is important. Similarly, though we've taken steps to randomize ballot order, we

may nevertheless be communicating implicit testimony to individuals that some candidates are better than others simply by the order in which we list them.

While I know of no direct evidence that this is how candidate order is understood by voters, there is indirect evidence. This evidence is best approached by considering other nudges. We've already mentioned the nudges that tend to feature as examples in the literature: the way in which changing the visual accessibility of food in the cafeteria line changes consumption patterns (Bucher et al. 2016) and changing the default option to a higher rate on an employment contract increases savings (Smith et al. 2013). Both of these nudges can be understood as providing implicit recommendations to agents.

The use of defaults to change behaviors has widely been seen as taking advantage of our cognitive laziness. But there's evidence that defaults are understood as communications (Fisher 2020a, 2020b; Levy 2019b). Agents tend to see default options as authoritatively recommended to them: both experimental work (McKenzie et al. 2006) and modeling (Carlin et al. 2013) suggests that ordinary people see defaults as reflecting expert opinion, and they change their attitudes to the default accordingly. The presentation of defaults is likely understood as communicative because it *is* communicative: the selection of defaults is typically meant (implicitly) to convey a recommendation. That's how the framing of options works: for example, a research and development team is more likely to be described in terms of success rate, rather than its failure rate, if the person thinks highly of it (Sher & McKenzie 2006). Framing of options is intended as communicative and understood as such. It's likely that selection of defaults functions the same way.

The selection of a default is the provision of implicit testimony: *this* option is best, or at least sufficiently good to be choiceworthy (isn't this exactly what we'd expect pretheoretically: wouldn't you be extremely surprised if a default option on a form or an employment contract wasn't a reasonable option? Imagine if your employment contract had as the default option saving 98 percent of your $60,000 salary into a retirement fund. Wouldn't you think whoever drew up the contract was incompetent?) Ballot order effects and the effects of making items more or less accessible or salient (by, for example, placing them at eye level) may be

understood in precisely the same way. While it's true that candidate quality doesn't, as a matter of fact, correlate with ballot order (except by chance), it may nevertheless be true that being guided by ballot order is a rational response to an implicit recommendation.

This account of how (canonical) nudges function to guide behavior doesn't merely provide an explanation of how they work (albeit an explanation falling well short of a full mechanistic account); it also *rationalizes* them. That is, it shows that and how it's rational to be guided by a nudge. It's rational to be guided by a nudge because it's rational to give due weight to a recommendation, implicit or not. A recommendation is higher-order evidence that an option is choiceworthy, and higher-order evidence is genuine evidence. Let's see how this works.

Framing effects are often regarded as paradigmatically irrational, on the grounds that how identical options are framed has a significant effect on whether they're preferred (e.g., Shafir and LeBoeuf 2002). Since it's the frame, and not the content of the option, that has altered, a change in preference is thought to be irrational: if your preference for A over B flips in response to a change that is irrelevant to which is better, than your preference flip is irrational. But how options are framed *isn't* irrelevant to which is better—far from it. While the first-order evidence is fixed across frames, the higher-order evidence is not: framing just is changing the higher-order evidence. *Of course* it's rational to be guided by a recommendation!

After all, no one ever objected to guidance by testimony on grounds like these. No one ever said "it's irrational to go to restaurant A rather than B just because your friend recommended it. After all, had she recommended B instead, that's where you'd have gone. Her recommendation doesn't change the options themselves." That's just how recommendations are *supposed* to work. They're supposed to provide higher-order evidence. They're guides to what the first-order facts are and therefore it's no objection to them that they leave these facts unaltered.

To be sure, it would be worrying if nudges were *compelling* causes of behavior: if they overrode other and better sources of information, for example. Recommendations don't overwhelm our better judgment and compel us to act. We respond judiciously to them: we integrate them

with other information available to us, including information about the person providing the recommendation. We probably won't go to Tofu Hut on our friend's recommendation if we know it's just failed a health inspection or we know she's very unreliable on questions like these. We take recommendations into account alongside other information and they are decisive for us only when we lack better information. The evidence strongly suggests that nudges work in the same way: they provide information we take into account, not mechanical shoves. The ballot order effect, for instance, may provide implicit testimony to everyone, but it makes a difference to the choices of only two groups of agents: those who have no real preference between candidates and those who have little information about them (Pasek et al. 2014). Ballot order effects influence behavior in just the way, and to the same degree, as testimony from a source regarded as somewhat reliable does, making a difference for those who have little else to go on. That's because ballot order effects *are* implicit testimony.

Nudges are sometimes defended on the grounds that it's *ecologically rational* to be guided by them (Gigerenzer 2015).[4] Use of a heuristic or bias is ecologically rational just in case it enables us reliably to get the right answer. *How* we get to that answer is irrelevant to its ecological rationality. In principle, ecological rationality can vindicate anything, if it turns out to work. If a benevolent God ensured that throwing dice was a reliable way to make investment decisions, it would be ecologically rational to rely on the dice. My defense of nudging doesn't depend on the notion of ecological rationality. Framing options, the selection of defaults and other ways of making options salient are ways of providing implicit testimony, and it's *directly* rational to be guided by testimony. Implicit testimony isn't merely correlated with the right answer, in the way in which (say) being banded might correlate with being venomous in snakes. Rather, implicit testimony is *evidence* for the right answer, and in being guided by it appropriately—in giving it the weight in our

[4] Schmidt (2019) defends the rationality of nudges on grounds like these. On his view, nudges are content rational—they enable us to achieve epistemic ends—but they sacrifice a great deal of process rationality by ignoring or failing to track the features that make an option choiceworthy. In contrast, I claim that nudges are process rational in an entirely orthodox manner. Nudges provide evidence for all agents. Their influence is directly rational, not (merely) ecologically rational.

cognition that reflects its actual evidential value as testimony—we're being guided by the evidence. Perhaps there are nudges that are merely ecologically rational, but most are (also) directly rational and do not bypass rational cognition at all.[5]

Those few people sympathetic to the thought that nudging might be rational see nudges as working through rational and non-rational channels simultaneously. One piece of evidence that's been cited for the conclusion that nudge-style influences work non-rationally is that their power increases when we're under cognitive load (when processing resources are scarce, for instance because the person is required to multi-task, or is fatigued or stressed; see, e.g Gilbert and Osborne, 1989; Krull and Erickson, 1995). This fact has been taken to indicate that they work (in part at least) by taking advantage of cognitive laziness or the fact that it is temporarily too difficult for us to make a decision. Ansher et al. (2014) make the point explicit: changing defaults at once provides recommendations to agents *and* takes advantage of non-rational dispositions.

The thought seems to be that if defaults brought about their effects through rational processes alone, we wouldn't see their influence increase under load. If they provided only rational inputs, then the person should respond to them when her processing capacity is undiminished just as strongly as when it is depleted. But that's a mistake. There is nothing irrational about putting more weight on testimony when we lack the resources to assess a claim for ourselves. We all accept that I ought to place more weight on your testimony when you're more expert in the relevant domain than I am. We might think of load analogously: while I'm under load, I should give greater weight to testimony because I'm temporarily less expert.

Ashner et al. cite another piece of evidence in arguing that default effects are partially irrational. The evidence comes from a study in which pulmonologists were asked whether they would prescribe a CT scan for

[5] It's difficult to come up with an example of a nudge that might be ecologically rational but is not directly rational. Perhaps priming behavior might be an example; whether primes should be understood as presenting us with evidence is difficult to assess. It's important to note that the priming literature has been a principal victim of the replication crisis in psychology: priming is (to my mind) real, but is a fragile and very weak influence on behavior.

a patient. In the control condition, 54 percent of pulmonologists ordered the scan. That establishes a baseline: given the symptoms described, roughly half will think a scan is warranted. In the other condition, participants were told that a scan had already been ordered but not yet performed. In this condition, only 29 percent of physicians cancelled the scan. Following the original authors of this study, Ansher et al. suggest that a mere (non-rational) bias drives the difference between conditions: "clinical information should dictate whether or not a CT scan should be performed […] whether or not it has been ordered or discontinued by the emergency department physician should be irrelevant" (Aberegg et al., 2005: 1499). Of course, that's false. Clinical information provides first-order evidence, and that evidence is obviously of critical importance to clinicians. But the attitudes of our epistemic peers—here represented by the decisions of other physicians—provides us with higher-order evidence, and when a question is difficult to settle (as the fact that pulmonologists split on whether to order the scan indicates), such evidence *should* be given significant weight. Again, the best explanation of how this information guides behavior is via the provision of higher-order evidence, and higher-order is genuine evidence.

Disagreement provides evidence about how well we have responded to our first-order evidence; it provides evidence we may have made a mistake in responding to it. Nudges and the like may not provide evidence like that (though they may). A recommendation, implicit or not, provides evidence about some other agent's attitude to an option. It may entirely replace first-order evidence. A recommendation may lead me to choose an option about which I know nothing; that is, about which I lack *any* first-order evidence. It provides evidence not about the facts that make an option choiceworthy, but about its choiceworthiness itself. It's not evidence about *my* evidence, but it's evidence about what the evidence, properly understood, supports.

In Praise of Nudges

Most (if not all) nudges provide agents with higher-order evidence. They are understood, implicitly, as encoding testimony. They're understood

in this kind of way because they *do* provide testimony: to nudge someone in the direction of an option is to recommend it to her. Just as agents frame communications to convey their opinions, so we make options salient or highlight them in other ways in order to recommend them. Giving greater weight to an option than we otherwise would just because it has been made salient to us in this kind of way is rational, because it's rational to give weight to recommendations (unless we have countervailing evidence, or reasons to distrust the person providing them). This remains true even holding first-order evidence fixed. That's just how testimony works and is supposed to work.

Of course, nudges (or their naturally or stochastically occurring analogues) may not present us with *reliable* testimony. Candidate order doesn't in fact correlate with candidate quality. The options may have been framed by agents who know little about the costs and benefits, or who seek to manipulate us to their own ends. But these facts are no objection to the claim that nudges work by offering us implicit testimony, nor to the claim that giving them due weight in our choices is rational. Exactly the same points, after all, apply to *explicit* testimony. Agents may offer testimony inadvertently (perhaps unaware that we're listening). They can offer testimony despite knowing little about the options, and I don't need to point out that agents may offer testimony in order to manipulate us to their own ends: that's obviously a central tactic of salespeople and advertisers.

Once we see that (most) nudges work by offering implicit testimony to agents, we're in a good position to see that many of their opponents have got things completely backwards. They demand, in effect, that we leave things as they are so that people are offered misleading testimony,[6] rather than change the context of choice so that people are offered testimony that genuinely tracks option quality. Though they don't recognize it, they're advocating the deception of others, rather than taking steps to ensure that they're told the truth. That's not respectful of agency: quite the opposite. Nudging well is offering honest testimony, and refusing to nudge is refusing to ensure that bad testimony is no longer offered.

[6] In the event the testimony isn't misleading, of course, there's no dispute: no one advocates intervening to change it.

There are nudges that don't appear to work through this kind of mechanism, but most nudges that have been proposed for the improvement of choice provide reliable higher-order evidence and thereby testimony. These nudges are defensible both on the grounds that they provide good evidence to agents, *and* on the grounds that they enhance individual or collective welfare. Some nudges may work through the provision of *unreliable* (first-order) evidence. For instance, fake potholes or fake speed bumps (in both cases painted on the road surface) have been used to slow traffic (Hamill 2008). I take no stand on whether such nudges are defensible; I note only that if they're indefensible it won't be on the ground that they bypass rational cognition. Deliberately misleading evidence is still evidence.[7]

Canonical nudges, however—and, as we'll soon see, other ways of engineering the epistemic environment—don't suffer from this problem. They're properly respectful of agency, because they work by providing the agent with reliable (usually higher-order) evidence and *thereby* improve our welfare. We may utilize such nudges in good conscience.

Stepping Back

In previous chapters, I suggested we needed to go beyond removing pollutants from the epistemic environment, and actually structure the environment to nudge agents toward better beliefs. I've suggested, for example that we should ensure that markers of expertise correlate with genuine expertise; that a view is represented in the media in rough correlation with the proportion of experts who hold it, and so on. Ensuring that higher-order evidence is reliable suggests policies that may be contentious: for instance, it suggests that news organizations shouldn't seek to balance competing experts when one view is very much better supported than another, because balancing speakers falsely conveys the impression that there's no consensus.

[7] Nudges like this are paradigm cases of paternalistic interventions: they're noble lies, intended to bring people to behave better by first getting them to believe falsehoods. They're defensible if (or when) such lies are defensible.

Elsewhere, I've suggested that a policy of ensuring that higher-order evidence is conveyed appropriately may support no-platforming certain speakers, on the grounds that provision of a platform itself provides higher-order evidence in favor of a view (Levy 2019c). While there are surely grounds for worry about some of these proposals (we may legitimately worry who can be appropriately trusted to implement nudges, about the potential for their misuse, and about restrictions on free speech), recognizing that these proposals aim at the provision of higher-order evidence disarms some objections. They shouldn't be seen as manipulative or disrespectful of agency: on the contrary, they're maximally respectful, in just the same way and for just the same reasons as giving people arguments and (first-order) evidence is maximally respectful of agency.

Throughout this book, I've argued that we're deeply social agents, agents who owe our epistemic success to the division of epistemic labor and the ways in which we scaffold cognition. Much of the scaffolding we rely on involves the flexible use of higher-order evidence, I now emphasize. Our use of social referencing—our use of cues as to what others believe to form our own beliefs—the conformity bias and the prestige bias, our outsourcing of belief to the environment and our reliance on distributed networks of agents and artifacts; all of these should be seen as reliance on higher-order evidence. Evidence about what the majority believes is higher-order evidence, as we saw in our discussion of how the numbers count when it comes to peer agreement and disagreement. The prestige bias consists in the use of indirect evidence—higher-order evidence—that certain ways of behaving bring success. That someone is prestigious is higher-order evidence that they behave, or think, well. We make certain facts salient to one another—sometimes through the design of the physical environment—to recommend them (and sometimes to provide implicit warnings, which is another form higher-order evidence can take). Peer review is, in part, the institutionalization of higher-order evidence: that a paper has received its imprimatur is (some) evidence in favor of its quality and its reliability.

If we engage in the kinds of strategies I recommend—nudging better belief—we won't be doing anything new. We've been nudging better belief forever: our epistemic success has *always* been dependent on

ensuring that higher-order evidence is reliable. Higher-order evidence is the real secret of our success. Correlatively, epistemic engineering is not dependent on our biases (understood as ways in which we fall short of rationality), or our (putative) cognitive laziness or even our bounded rationality. Rather, it takes advantages of our rational faculties.

Concluding Thoughts

Rational Animals After All

The Enlightenment celebrated the power of human reason. In more recent times, a number of psychologists and naturalistic philosophers have attempted to replace the Enlightenment conception of our cognitive powers with a deflationary and allegedly more realistic picture. On this picture, often presented within the framework of dual process theory, the kind of rationality the Enlightenment highlighted, impressive as it genuinely is, is only a small part of the fuller picture. Type 2 cognition is a scarce resource, and we must be selective in deploying it. The rest of the time—most of the time—we must rely on fast and frugal Type 1 cognition (Bargh & Chartrand 1999). Even when we do deploy Type 2 cognition, we remain reliant on Type 1 cognition for inputs, which limits the power of Type 2 cognition to correct errors. This is bad news for our rationality, because Type 1 cognition is inflexible and sometimes atavistic.

Pushback against the claim that our cognition is lazy, inflexible and pervasively irrational has come mainly from advocates of ecological conceptions of rationality. They emphasize how well adapted such cognition is to the threats we actually faced in the environment of evolutionary adaptiveness. They concede, of course, that we pay costs for relying on it: for example, we respond to stimuli as if they were threats while knowing that they're harmless. Even when we deliberate carefully, we may be subject to prejudices we disavow, because the reasons we consciously process have already been assigned weights prior to deliberation (Uhlmann & Cohen 2005). But these are costs a well-designed thinker would pay, proponents of ecological rationality maintain. It's better to respond quickly to a possible threat then to deliberate longer:

Bad Beliefs: Why They Happen to Good People. Neil Levy, Oxford University Press. © Neil Levy 2022.
DOI: 10.1093/oso/9780192895325.003.0007

we'd do better to pay the cost of a false positive then to run the risk of being bitten or mauled. If we're irrational due to our reliance on fast and frugal cognition, we're rational to be irrational in this way.

In this book, I've presented a very different picture. I've suggested that we're more rational than naturalistic philosophers have tended to think. We've failed to see how rational we are because we've been looking for rationality in all the wrong places. We've been looking at individual cognition and at first-order evidence to vindicate our conception of ourselves as rational agents. Both of these things matter, of course: they matter a great deal. But apparent failures to rely on them often don't indicate departures from rationality. They indicate a rational outsourcing of our cognition, a reliance on the division of epistemic labor, and the appropriate use of higher-order evidence.

With this picture in place, we can now see that we're *individually* more rational than we sometimes seem, though our individual rationality doesn't take the form we expect. The behavior of other agents is higher-order evidence for us, and individually we respond to it appropriately. Deference to experts is an appropriate use of higher-order evidence; so is the use of the conformity bias and the prestige bias. The use of environmental cues is the use of higher-order evidence: it renders options salient to us. Most of the time, we respond appropriately to the communicative cues in our environments. Our individual rationality doesn't consist in our processing of first-order evidence alone. It consists, also and importantly, in the use of higher-order evidence, made available to us through distributed and outsourced cognition.

Our flexible and intelligent response to higher-order evidence is rational, whether it's the product of conscious deliberation or of automatic processing. Of course, deliberation (conscious or not) about first-order evidence is also essential to our epistemic success. My deference to experts is appropriate because *they* have produced a body of knowledge, and their knowledge production is heavily reliant on first-order evidence. But no significant epistemic achievement is the product of deliberation about first-order evidence alone. Even scientists must defer, even on their own terrain, because no individual is able to grasp all the methods and tools and data she must nevertheless rely on for her work. Within the narrow sphere of our expertise, our reliance on first-order

evidence is relatively heavy; elsewhere, higher-order evidence plays a much greater role.

At this point, a worry about self-defeat obviously arises. This book is itself the product of individual deliberation, and argues against an orthodoxy that emphasizes first-order evidence, intellectual autonomy and individual deliberation. If I'm confident that I'm able to see the problems in the orthodox view, then I seem committed to thinking that individual cognition can successfully strike out on its own, contrary to my own claims. Aren't I committed to making an exception of myself, in a way that is at best unprincipled? Deference for thou; not for me.

There is, I acknowledge, a tension between the message of this book and the very act of advocating that message. To some degree, this tension is endemic to philosophy itself, insofar as it tries to show "how things in the broadest possible sense of the term hang together in the broadest possible sense of the term" (Sellars 1962). This ambition requires stepping back and trying to find what is common to disparate areas of inquiry. In an age of hyperspecialists, philosophy still pretends to the role of the generalist. Inevitably, the generalist runs the risk of a failure of understanding when they take as their subject matter the findings or the nature of specialized disciplines. Many philosophers, including me, have responded to this kind of worry by backing away from generalist pretensions and instead limiting ourselves to more specific domains, where we can hope to make headway. But this book attempts much more general claims about human knowledge and its acquisition and thereby runs full tilt into these concerns, and in a particularly pointed way due to its advocacy of deference.

I take comfort from the fact that I'm not entirely out on my own here. In this book, I'm leveraging and building on the work of many others: work in cultural evolution, in psychology, in social epistemology, and in other fields. To a large degree, I *am* deferring to these thinkers (moreover one of my major aims, albeit one that's in the background, has been to make us more accepting of a science—climate change—that I lack any real capacity to understand). I do take my own advice in this regard, if not as extensively as perhaps I should. Nevertheless, to some degree I am caught in worries about self-contradiction. I present what I take to be first-order evidence in favor of relying on higher-order evidence,

arguments in favor of deference at odds with an orthodoxy that denigrates it.

Of course, I haven't advocated relying on higher-order evidence *rather than* first-order evidence. I've stressed that higher-order is reliable, in very important part, because it's generated by people who are grappling with the first-order evidence. All cognition is dependent on both first-order and higher-order evidence, to varying degrees in different contexts, depending on our capacity to assess each kind of evidence. To some extent, that fact mitigates the tension. Still, I can't complain to have the kind of expertise that would warrant me in setting aside the near-universal consensus in favor of individual reasoning and striking out on my own. The tension remains unresolved. Here I can only appeal to you for help. Come over to my side. Once there are enough of us, I can comfortably advocate deferring to the new orthodoxy. If you're on the fence about my arguments, I hope concern for my comfort will tip you over into accepting them.

Let me finish with a few words about the Enlightenment. Must we abandon its legacy, if we accept the picture of ourselves I've urged here? Perhaps not. *Sapere Aude!*, Kant's injunction, is most naturally interpreted individualistically. Kant calls on us to emerge from "immaturity," which he characterizes as "the inability to use one's own understanding without the guidance of others." He thus calls on us to use our "*own* understanding" (Kant, 1991: 54; emphasis in original). I, too, advocate we use our own understanding. On my picture, though, there is no conflict between such use and apt deference. We should *not* use our understanding without the guidance of others; instead, a primary function of our understanding is in orienting us well toward such guidance.

Kant called on us to change our epistemic strategies, to rely more on our individual judgment and less on the judgments of others. Insofar as I have advice for each of us, as individuals, it's to rely on others more and better (and of course, insofar as we're able, to engineer the epistemic environment to support such reliance). We err in overemphasizing individualism, not in deferring too much. Does that entail abandoning the legacy of the Enlightenment? Not necessarily.

First, perhaps Kant was right to call on his contemporaries to think more for themselves and to defer less. It is no part of the picture

I've presented here to claim that individual deliberation over first-order evidence isn't a central component of our cognitive success. I've argued we ought to defer to the scientists, and that scientists must defer to one another. But deference has its limits. We should defer to scientists in very important part because they've deployed their individual cognition in the domain of their expertise. Admittedly, they've deployed it in a way that is socially and institutionally supported, and in ways that are heavily imbricated with deference; they've employed it nevertheless, and their reliability is partly (but only partly) due to that fact. In Kant's historical context, perhaps people had too little opportunity, or too little motivation, to deploy their individual cognition in these ways (I leave that as a question for historians).

Second, we needn't see the legacy of the Enlightenment as exhausted by this heavy emphasis on *individual* rationality, in what Kant regards as its mature form ("without the guidance of others"). Rationality, in its fullest sense, is, roughly, the deployment of cognition in the effective service of truth by appropriate response to the evidential content of information. Many psychologists and philosophers see us as rationally irrational: we deploy our cognition in the effective service of truth but we do so through the use of heuristics and other fast and frugal processes that do *not* respond appropriately to the content of our information. We respond irrationally—in ways that are not warranted by our evidence—but we're rational to do so. On my account, we are rationally rational. We respond to the *higher-order* evidence encoded in our environment and in the assertions of others, by deferring to them or even self-attributing beliefs. We do so in the service of truth. We're rational animals after all, even if our rationality is somewhat different to how we imagined it. We need to have the courage to use one another's understanding as well as our own.

References

Aberegg, S. K., Haponik, E. F., & Terry, P. B. (2005). "Omission bias and decision making in pulmonary and critical care medicine," *Chest*, 128/3: 1497–505.

Acerbi, A., & Mesoudi, A. (2015). "If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution," *Biology & Philosophy*, 30/4: 481–503.

Ackermann, R., & DeRubeis, R. J. (1991). "Is depressive realism real?," *Clinical Psychology Review*, 11/5: 565–84.

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge: Cambridge University Press.

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (2017). "Development and validation of a multi-dimensional measure of intellectual humility," *PLoS ONE*, 12/8. DOI: 10.1371/journal.pone.0182950

Allcott, H., & Gentzkow, M. (2017). "Social media and fake news in the 2016 Election," *Journal of Economic Perspectives*, 31/2: 211–36.

Anderson, E. (2011). "Democracy, public policy, and lay assessments of scientific testimony," *Episteme*, 8/2: 144–64.

Anderson, L. R., & Holt, C. A. (1997). "Information cascades in the laboratory," *American Economic Review*, 87: 847–62.

Ansher, C., Ariely, D., Nagler, A., Rudd, M., Schwartz, J., & Shah, A. (2014). "Better medicine by default," *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 34/2: 147–58.

Associated Press. (2020). "In Iran, false belief a poison fights coronavirus kills hundreds." *Snopes.com*. Retrieved April 5, 2020, from <https://www.snopes.com/ap/2020/03/27/in-iran-false-belief-a-poison-fights-coronavirus-kills-hundreds/>

Astuti, R., & Harris, P. (2008). "Understanding mortality and the life of the ancestors in rural Madagascar," *Cognitive Science*, 32: 713–40.

Atran, S. (2002). *In Gods We Trust: The Evolutionary Landscape of Religion*., New edition. Oxford: Oxford University Press.

Bada, J. L., & Lazcano, A. (2009). "The origin of life." In Ruse, M. & Travis, J. (eds.), *Evolution: The First Four Billion Years*, pp. 49–79. Cambridge, Mass: Harvard University Press.

Ballantyne, N. (2019). *Knowing Our Limits.* Oxford: Oxford University Press.

Banerjee, N., Song, L., & Hasemyer, D. (2015). "Exxon's own research confirmed fossil fuels' role in global warming decades ago." *Inside Climate News*. Retrieved November 24, 2019, from <http://insideclimatenews.org/news/15092015/exxons-own-research-confirmed-fossil-fuels-role-in-global-warming>

Barber, M., & Pope, J. C. (2019). "Does party trump ideology? Disentangling party and ideology in America," *American Political Science Review*, 113/1: 38–54.

Bardon, A. (2019). *The Truth about Denial: Bias and Self-Deception in Science, Politics, and Religion*. Oxford and New York: Oxford University Press.

Bargh, J. A., & Chartrand, T. L. (1999). "The unbearable automaticity of being," *American Psychologist*, 54/7: 462–79.

Barrett, J. L. (1999). "Theological correctness: Cognitive constraint and the study of religion," *Method & Theory in the Study of Religion*, 11/4: 325–39.

Barrett, J. L. (2001). "How ordinary cognition informs petitionary prayer," *Journal of Cognition and Culture*, 1/3: 259–69.

Bayne, T. J., & Pacherie, E. (2005). "In defence of the doxastic conception of delusions," *Mind and Language*, 20/2: 163–88.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford, New York: Oxford University Press.

Berinsky, A. J. (2018). "Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding," *The Journal of Politics*, 80/1: 211–24.

Bishop, M. A., & Trout, J. D. (2004). *Epistemology and the Psychology of Human Judgment*. Oxford: Oxford University Press.

Blake, A. (2016). "'Pizzagate' theory believed by nearly half of Republicans: Poll", *Washington Times*, December 28.

Blancke, S., Boudry, M., & Pigliucci, M. (2017). "Why do irrational beliefs mimic science? The cultural evolution of pseudoscience," *Theoria*, 83/1: 78–97.

Bloudoff-Indelicato, M. (2015). "Backlash after Frontiers journals added to list of questionable publishers," *Nature News*, 526/7575: 613.

Booker, C. (2009). "Climate change: This is the worst scientific scandal of our generation." *The Telegraph*, November 28.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2010). "Salience theory of choice under risk," *The Quarterly Journal of Economics*, 127/2: 1243–1285.

Borick, C. P., & Rabe, B. G. (2017). "Personal experience, extreme weather events, and perceptions of climate change," *Oxford Research Encyclopedia of Climate Science*. DOI: 10.1093/acrefore/9780190228620.013.311

Boudry, M. (2018). "Replicate after reading: On the extraction and evocation of cultural information," *Biology & Philosophy*, 33/3: 27.

Bovens, L. (2008). "The ethics of nudge." In Hansson, M. J. & Grüne-Yanoff, T. (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology.*, pp. 207–20. Berlin: Springer, Theory and Decision Library A.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). "The cultural niche: Why social learning is essential for human adaptation," *Proceedings of the National Academy of Sciences*, 108/Supplement 2: 10918–25.

Boyer, P. (2000). "Functional origins of religious concepts: Ontological and strategic selection in evolved minds," *The Journal of the Royal Anthropological Institute*, 6/2: 195–214.

Brennan, D. (2018). "One-third of Americans don't believe 6 million Jews were murdered during the Holocaust." *Newsweek*, December 4. Retrieved December 2, 2019, from <https://www.newsweek.com/one-third-americans-dont-believe-6-million-jews-were-murdered-during-holocaust-883513>

Brennan, Jason. (2016). *Against Democracy*. Princeton: Princeton University Press.

Brennan, Johnny. (2020). "Can novices trust themselves to choose trustworthy experts? Reasons for (reserved) optimism," *Social Epistemology*, 34/3: 227–40.

Broniatowski, D. A., Jamison, A. M., Qi, S., Al Kulaib, L., Chen, T., Benton, A., Quinn, S. C., et al. (2018). "Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate," *American Journal of Public Health*, 108/10: 1378–84.

Brooks, R. A. (1990). "Elephants don't play chess," *Robotics and Autonomous Systems*, Designing Autonomous Agents, 6/1: 3–15.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). "The psychological impact of quarantine and how to reduce it: Rapid review of the evidence," *The Lancet*, 395/10227: 912–20.

Brownstein, M., Madva, A., & Gawronski, B. (2019). "What do implicit measures measure?," *WIREs Cognitive Science*, 10/5: e1501.

Bucher, T., Collins, C., Rollo, M. E., McCaffrey, T. A., Vlieger, N. D., Bend, D. V. der, Truby, H., et al. (2016). "Nudging consumers towards healthier choices: A systematic review of positional influences on food choice," *British Journal of Nutrition*, 115/12: 2252–63.

Bullock, J. G., & Lenz, G. (2019). "Partisan bias in surveys," *Annual Review of Political Science*, 22/1: 325–42.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2015). "Partisan bias in factual beliefs about politics," *Quarterly Journal of Political Science*, 10/4: 519–78.

Burcham, P. C. (2008). "Toxicology down under: Past achievements, present realities, and future prospects," *Chemical Research in Toxicology*, 21/5: 967–70.

Buskell, A. (2016). "Cultural longevity: Morin on cultural lineages," *Biology & Philosophy*, 31/3: 435–46.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., et al. (2018). "Evaluating the replicability of social science experiments in nature and science between 2010 and 2015," *Nature Human Behaviour*, 2/9: 637.

Capstick, S., Whitmarsh, L., Poortinga, W., Pidgeon, N., & Upham, P. (2015). "International trends in public perceptions of climate change over the past quarter century," *Wiley Interdisciplinary Reviews: Climate Change*, 6/1: 35–61.

Carl, N., Richards, L., & Heath, A. (2019). "Leave and Remain voters' knowledge of the EU after the referendum of 2016," *Electoral Studies*, 57: 90–8.

Carlin, B. I., Gervais, S., & Manso, G. (2013). "Libertarian paternalism, information production, and financial decision making," *The Review of Financial Studies*, 26/9: 2204–28.

Carruthers, P. (2013). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Carter, J. A., & McKenna, R. (2020). "Skepticism motivated: On the skeptical import of motivated reasoning," *Canadian Journal of Philosophy*, 50/6: 702–718.

Cassam, Q. (2018). *Vices of the Mind: From the Intellectual to the Political*. Oxford University Press.

Castelvecchi, D. (2015). "Physics paper sets record with more than 5,000 authors," *Nature News*. DOI: 10.1038/nature.2015.17567

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The History and Geography of Human Genes*, Abridged edition. Princeton, NJ: Princeton University Press.

Chater, N. (2018). *The Mind Is Flat: The Illusion of Mental Depth and the Improvised Mind*. New Haven, CT: Yale University Press.

Chawla, A. D. S. (2016). "Journal reverses acceptance of study linking vaccines to autism." *Retraction Watch*, December 9. Retrieved January 26, 2020, from <https://retractionwatch.com/2016/12/09/journal-reverses-acceptance-study-linking-vaccines-autism/>

Chiappori, P.-A., Levitt, S., & Groseclose, T. (2002). "Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer," *American Economic Review*, 92/4: 1138–51.

Christensen, D. (2007). "Epistemology of disagreement: The good news," *The Philosophical Review*, 116/2: 187–217.

Chudek, M., Heller, S., Birch, S., & Henrich, J. (2012). "Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning," *Evolution and Human Behavior*, 33/1: 46–56.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.

Clark, A., & Chalmers, D. J. (1998). "The extended mind," *Analysis*, 58/1: 7–19.

Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford: Clarendon Press.

Coady, D. (2006). "When experts disagree," *Episteme*, 3/1–2: 68–79.

Coady, D. (2007). "Are conspiracy theorists irrational?" *Episteme*, 4/2: 193–204.

Coady, D., & Corry, R. (2013). *The Climate Change Debate: An Epistemic and Ethical Enquiry*, 2013 edition. London: Palgrave Pivot.

Cohen, G. L. (2003). "Party over policy: The dominating impact of group influence on political beliefs," *Journal of Personality and Social Psychology*, 85/5: 808–22.

Cohn, R. (2013). "Bob Inglis: The Republican who believes in climate change." *The Guardian*, February 15.

Conly, S. (2013). *Against Autonomy: Justifying Coercive Paternalism*. Cambridge University Press.

Connor, P., Sullivan, E., Alfano, M. & Tintarev, N. (forthcoming). "Motivated numeracy and active reasoning in a Western European Sample," *Behavioural Public Policy*. doi: https://doi.org/10.1017/bpp.2020.32

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., et al. (2016). "Consensus on consensus: a synthesis of consensus estimates on human-caused global warming," *Environmental Research Letters*, 11/4: 048002.

Cooper, J. (2007). *Cognitive Dissonance: 50 Years of a Classic Theory*. SAGE Publications.

Cosmides, L., & Tooby, J. (1992). "Cognitive adaptations for social exchange." In Barkow, J. H., Tooby, J., & Cosmides, L. (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pp. 163–228. Oxford and New York: Oxford University Press.

Crone, D. L., & Levy, N. (2019). "Are free will believers nicer people? (Four studies suggest not) – Damien L. Crone, Neil L. Levy, 2019," *Social Psychological and Personality Science*, 10/5: 612–19.

Darcy, R., & McAllister, I. (1990). "Ballot position effects," *Electoral Studies*, 9/1: 5–17.

Dawkins, R. (1989). *The Selfish Gene*. 2nd revised edition. Oxford and New York: Oxford Paperbacks.

DeCruz, H. (2020). "Believing to belong: Addressing the novice-expert problem in polarized scientific communication," *Social Epistemology*, 34/5: 440–52.

Dennett, D. C. (1992). *Consciousness Explained*. 1st edition. Boston: Back Bay Books.

Ditto, P., & Lopez, D. (1992). "Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions," *Journal of Personality and Social Psychology*, 63: 568–84.

Dobzhansky, T. (1973). "Nothing in biology makes sense except in the light of evolution," *The American Biology Teacher*, 35/3: 125–9.

Doherty, C., Kiley, J., & Johnson, B. (2017). *Sharp Partisan Divisions in Views of National Institutions*., p. 18. Pew Research Center.

Doris, J. M. (2015). *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press.

Doris, J. M. (2018). "Précis of talking to our selves: Reflection, ignorance, and agency," *Behavioral and Brain Sciences*, 41: 1–12.

Drummond, C., & Fischhoff, B. (2017). "Individuals with greater science literacy and education have more polarized beliefs on controversial science topics," *Proceedings of the National Academy of Sciences*, 114/36: 9587–92.

Duhaime, E. P. (2015). "Is the call to prayer a call to cooperate? A field experiment on the impact of religious salience on prosocial behavior," *Judgment and Decision Making*, 10/6: 4.

Dunning, D. (2019). "Gullible to ourselves." In Forgas, M. & Baumeister, R. (eds) *The Social Psychology of Gullibility: Conspiracy Theories, Fake News and Irrational Beliefs*, pp. 217–33. Routledge: Abingdon.

Earp, B. (2016). "What did the OSC replication initiative reveal about the crisis in psychology?," *BMC Psychology*, 4: 1–19.

Edelman, B. (2009). "Markets: Red light states: Who buys online adult entertainment?," *Journal of Economic Perspectives*, 23: 209–20.

Elga, A. (2007). "Reflection and disagreement," *Noûs*, 41: 478–502.

Environmental and Energy Study Institute. (2019). "Fact sheet: Fossil fuel subsidies: A closer look at tax breaks and societal costs" Retrieved November 23, 2019, from <https://www.eesi.org/papers/view/fact-sheet-fossil-fuel-subsidies-a-closer-look-at-tax-breaks-and-societal-costs>

Evans, G. (1982). *The Varieties of Reference*. Oxford and New York: Oxford University Press.

Evans, J. St. B. T. (2008). "Dual-processing accounts of reasoning, judgment, and social cognition," *Annual Review of Psychology*, 59: 255–78.

Evans, J. St. B. T., & Stanovich, K. E. (2013). "Dual-process theories of higher cognition: advancing the debate," *Perspectives on Psychological Science*, 8/3: 223–41. DOI: 10.1177/1745691612460685

Everett, J. A. C., & Earp, B. D. (2015). "A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers," *Frontiers in Psychology*, 6. DOI: 10.3389/fpsyg.2015.01152

Every-Palmer, S., & Howick, J. (2014). "How evidence-based medicine is failing due to biased trials and selective publication," *Journal of Evaluation in Clinical Practice*, 20/6: 908–14.

Fantl, J. (2018). *The Limitations of the Open Mind*. Oxford: Oxford University Press.

Farrell, H. (2013). "Krauthammer was very, very wrong." *Washington Monthly—Politics*. April 22. Retrieved January 26, 2020, from <https://washingtonmonthly.com/2013/04/22/krauthammer-was-very-very-wrong/>

Ferguson, A. C. (2015). "Frontiers lets HIV denial article stand, reclassifies it as 'opinion.'" *Retraction Watch*. February 24. Retrieved January 26, 2020, from <https://retractionwatch.com/2015/02/24/frontiers-lets-hiv-denier-article-stand-reclassifies-it-as-opinion/>

Fine, C. (2013). "Is there neurosexism in functional neuroimaging investigations of sex differences?," *Neuroethics*, 6/2: 369–409.

Fisher, S. A. (2020a). "Meaning and framing: The semantic implications of psychological framing effects," *Inquiry*, 0/0: 1–24. Routledge.

Fisher, S. A. (2020b). "Rationalising framing effects: At least one task for empirically informed philosophy," *Crítica: Revista Hispanoamericana de Filosofía*, 52/156: 5–30.

Flynn, E., & Smith, K. (2012). "Investigating the mechanisms of cultural acquisition," *Social Psychology*, 43/4: 185–95.

Forstmann, M., & Sagioglou, C. (2020). "Religious concept activation attenuates cognitive dissonance reduction in free-choice and induced compliance paradigms," *The Journal of Social Psychology*, 160/1: 75–91.

Fridland, E. (2018). "Do as I say and as I do: Imitation, pedagogy, and cumulative culture," *Mind & Language*, 33/4: 355–77.

Fritz, J. (2018). "Conciliationism and moral spinelessness," *Episteme*, 15/1: 101–18.

Funk, C., & Kennedy, B. (2019). "How Americans see climate change in 5 charts." *Pew Research Center*. Retrieved November 24, 2019, from <https://www.pewresearch.org/fact-tank/2019/04/19/how-americans-see-climate-change-in-5-charts/>

Funkhouser, E. (forthcoming). "A tribal mind: Beliefs that signal group identity or commitment," *Mind and Language*. DOI: 10.1111/mila.12326

Galloway, J. N., Likens, G. E., & Hawley, M. E. (1984). "Acid precipitation: Natural versus anthropogenic components," *Science*, 226/4676: 829–31.

Gauchat, G. (2012). "Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010," *American Sociological Review*, 77/2: 167–87.

Gergely, G., Bekkering, H., & Király, I. (2002). "Rational imitation in preverbal infants," *Nature*, 415/6873: 755–755.

Gigerenzer, G. (2015). "On the supposed evidence for libertarian paternalism," *Review of Philosophy and Psychology*, 6/3: 361–83.

Gilbert, D., & Osborne, R. (1989). "Thinking backward: Some curable and incurable consequences of cognitive busyness," *Journal of Personality and Social Psychology*, 57/6: 940–9.

Goldberg, S. C. (2010). *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press

Goldman, A. I. (2001). "Experts: Which ones should you trust?," *Philosophy and Phenomenological Research*, 63/1: 85–110.

Goldstein, D. G., & Gigerenzer, G. (2002). "Models of ecological rationality: The recognition heuristic," *Psychological Review*, 109/1: 75–90. US: American Psychological Association.

Grant, B. (2009). "Elsevier published 6 fake journals." *The Scientist Magazine*, May 6. <https://www.the-scientist.com/the-nutshell/elsevier-published-6-fake-journals-44160>

Green, M. S. (1998). "Direct reference and implicature," *Philosophical Studies*, 91/1: 61–90.

Guerrero, A. (2017). "Living with ignorance in a world of experts." In *Perspectives on Ignorance from Moral and Social Philosophy*, pp. 168–97. Routledge: New York.

Guess, A., Nagler, J., & Tucker, J. (2019). "Less than you think: Prevalence and predictors of fake news dissemination on Facebook," *Science Advances*, 5/1: eaau4586. DOI: 10.1126/sciadv.aau4586

Hall, L., Johansson, P., & Strandberg, T. (2012). "Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey," *PLOS ONE*, 7/9: e45457. DOI: 10.1371/journal.pone.0045457

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., & Johansson, P. (2013). "How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions," *PloS One*, 8/4: e60554. DOI: 10.1371/journal.pone.0060554

Hamill, S. D. (2008). "To slow speeders, Philadelphia tries make-believe." *The New York Times*, July 12.

Hardwig, J. (1985). "Epistemic dependence," *Journal of Philosophy*, 82/7: 335–49.

Harris, K. (2018). "What's epistemically wrong with conspiracy theorising?," *Royal Institute of Philosophy Supplement*, 84: 235–57.

Harris, P. (2012). *Trusting What You're Told.* Cambridge, MA: Harvard University Press.

Harris, P. (2013). "One in four Americans think Obama may be the antichrist, survey says." *The Guardian*, April 3.

Harris, P., & Giménez, M. (2005). "Children's acceptance of conflicting testimony: The case of death," *Journal of Cognition and Culture*, 5: 143–64.

Hastorf, A. H., & Cantril, H. (1954). "They saw a game: A case study," *Journal of Abnormal Psychology*, 49/1: 129–34.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). "The extent and consequences of P-hacking in science," *PLOS Biology*, 13/3: e1002106. DOI: 10.1371/journal.pbio.1002106

Henrich, J. (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.

Henrich, J., & Boyd, R. (1998). "The evolution of conformist transmission and the emergence of between-group differences," *Evolution and Human Behavior*, 19/4: 215–41.

Henrich, J., & Gil-White, F. J. (2001). "The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission," *Evolution and Human Behavior*, 22/3: 165–96.

Henrich, J., & Henrich, N. (2010). "The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins," *Proceedings of the Royal Society B: Biological Sciences*, 277/1701: 3715–24.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). "The weirdest people in the world?," *Behavioral and Brain Sciences*, 33/2–3: 61–83.

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge, MA: Belknap Press.

Hornsey, M. J., Harris, E. A., Bain, P. G., & Fielding, K. S. (2016). "Meta-analyses of the determinants and outcomes of belief in climate change," *Nature Climate Change*, 6/6: 622–6.

Hudson, M. (2018). "George Bush Sr could have got in on the ground floor of climate action – history would have thanked him." *The Conversation*, December 6. Retrieved November 30, 2019, from <http://theconversation.com/george-bush-sr-could-have-got-in-on-the-ground-floor-of-climate-action-history-would-have-thanked-him-108050>

Ioannidis, J. P. A. (2020). "Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures," *European Journal of Clinical Investigation*, 50/n/4: e13222. DOI: 10.1111/eci.13222

IPCC. (2018). *The Special Report on Global Warming of 1.5 °C*. Retrieved December 20, 2019, from <https://www.ipcc.ch/sr15/>

Jerit, J., & Barabas, J. (2012). "Partisan perceptual bias and the information environment," *The Journal of Politics*, 74/3: 672–84.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). "Failure to detect mismatches between intention and outcome in a simple decision task," *Science (New York, N.Y.)*, 310/5745: 116–9.

John, L. K., Loewenstein, G., & Prelec, D. (2012). "Measuring the prevalence of questionable research practices with incentives for truth telling," *Psychological Science*, 23/5: 524–32.

Kahan, D. M. (2008). *Cultural Cognition as a Conception of the Cultural Theory of Risk* (SSRN Scholarly Paper No. ID 1123807). Rochester, NY: Social Science Research Network. Retrieved November 27, 2019, from <https://papers.ssrn.com/abstract=1123807>

Kahan, D. M. (2014). *Vaccine Risk Perceptions and* Ad Hoc *Risk Communication: An Empirical Assessment* (SSRN Scholarly Paper No. ID 2386034). Rochester, NY: Social Science Research Network. Retrieved November 30, 2019, from <https://papers.ssrn.com/abstract=2386034>

Kahan, D. M. (2015). "Climate-science communication and the measurement problem," *Political Psychology*, 36/s1: 1–43.

Kahan, D. M. (2017). *Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition* (SSRN Scholarly Paper No. ID 2973067). Rochester, NY: Social Science Research Network. Retrieved July 3, 2019, from <https://papers.ssrn.com/abstract=2973067>

Kahan, D. M., Hoffman, D. A., Braman, D., Evans Peterman, D., & Rachlinski, J. J.(2011). *"They Saw a Protest": Cognitive Illiberalism and the Speech-Conduct Distinction* (SSRN Scholarly Paper No. ID 1755706). Rochester, NY: Social Science Research Network. Retrieved November 26, 2019, from <https://papers.ssrn.com/abstract=1755706>

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2010). *Cultural Cognition of Scientific Consensus* (SSRN Scholarly Paper No. ID 1549444). Rochester, NY: Social Science Research Network. Retrieved March 11, 2019, from <https://papers.ssrn.com/abstract=1549444>

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). "Motivated numeracy and enlightened self-government," *Behavioural Public Policy*, 1/1: 54–86.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Penguin UK.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kant, I. (1991). "What Is Enlightenment?" *Political Writings*. Cambridge: Cambridge University Press.

Keil, F. C., & Wilson, R. A. (2000). "The Shadows and Shallows of Explanation." In Keil, F. C. & Wilson, R. A. (eds.), *Minds and Machines*, pp. 137–59. Cambridge, MA: MIT Press.

Keller, S. (2015). "Empathizing with scepticism about climate change." In Moss J. (ed.), *Climate Change and Justice*, pp. 219–35. Cambridge University Press: Cambridge.

Kennedy, B., & Hefferon, M. (2019). "U.S. concern about climate change is rising, but mainly among Democrats." *Pew Research Center*. Retrieved November 24, 2019, from <https://www.pewresearch.org/fact-tank/2019/08/28/u-s-concern-about-climate-change-is-rising-but-mainly-among-democrats/>

Kennedy, H. (2019). "I fact-checked Naomi Wolf's new book on gay rights – the mockery she has faced for one error disguises the real outrage we should feel." *The Independent*, June 28.

Kilov, D. (forthcoming). "The brittleness of expertise and why it matters," *Synthese*, 1–25. DOI: 10.1007/s11229-020-02940-5

King, A., & Leigh, A. (2009). "Are ballot order effects heterogeneous?," *Social Science Quarterly*, 90/1: 71–87.

Kitcher, P. (1987). *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge, MA: MIT Press.

Kripke, S. A. (2011). *Philosophical Troubles: Collected Papers*. Oxford: Oxford University Press.

Krug, E. C., & Frink, C. R. (1983). "Acid rain on acid soil: A new perspective," *Science*, 221/4610: 520–5.

Krull, D. S., & Erickson, D. J. (1995). "Judging situations: On the effortful process of taking dispositional information into account," *Social Cognition*, 13/4: 417–38.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kumar, V. (2016). "Nudges and bumps," *Georgetown Journal of Law and Public Policy*, 14: 861–76.

Kunda, Z. (1990). "The case for motivated reasoning," *Psychological Bulletin*, 108/3: 480–98.

Kunzru, H. (2020). "For the Lulz." *New York Review of Books*, March 26. <https://www.nybooks.com/articles/2020/03/26/trolls-4chan-gamergate-lulz/>

Kurtzleben, D. (2016). "POLL: White evangelicals have warmed to politicians who commit 'immoral' acts." *NPR.org*. Retrieved December 17, 2019, from <https://www.npr.org/2016/10/23/498890836/poll-white-evangelicals-have-warmed-to-politicians-who-commit-immoral-acts>

LaCapria, K. (2016). "Fact check: American pediatricians issue statement that transgenderism is 'child abuse.'" *Snopes.com*. Retrieved January 25, 2020, from <https://www.snopes.com/fact-check/americas-pediatricians-gender-kids/>

Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

Lackey, J. (2010). "A justificationist view of disagreement's epistemic significance." In Haddock, A., Millar, A., & Pritchard, D. (eds.), *Social Epistemology*, pp. 298–325. Oxford: Oxford University Press.

Lackey, J., & Sosa, E. (2006). *The Epistemology of Testimony*. Oxford: Oxford University Press.

Laland, K. N. (2017). *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton, NJ: Princeton University Press.

Langdon-Neuner, E. (2008). "Medical ghost-writing," *Mens Sana Monographs*, 6/1: 257–73.

Lawson, A. E., & Worsnop, W. A. (1992). "Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment," *Journal of Research in Science Teaching*, 29/2: 143–66.

Leeson, P. T. (2012). "Ordeals," *The Journal of Law and Economics*, 55/3: 691–714.

Leggett, N. C., Thomas, N. A., Loetscher, T., & Nicholls, M. E. R. (2013). "The life of p: 'Just significant' results are on the rise," *Quarterly Journal of Experimental Psychology*, 66/12: 2303–9.

Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Rosenthal, S. (2014). *Climate Change in the American Mind*. Yale Project on Climate Change Communication. New Haven, CT: Yale University and George Mason University.

Lerman, A. E., Sadin, M. L., & Trachtman, S. (2017). "Policy uptake as political behavior: Evidence from the affordable care act," *American Political Science Review*, 111/4: 755–70.

Levy, N. (2006). "Open-mindedness and the duty to gather evidence," *Public Affairs Quarterly*, 20/1: 55–66.

Levy, N. (2007). "Radically socialized knowledge and conspiracy theories," *Episteme*, 4/2: 181–92.

Levy, N. (2012). "Ecological engineering: Reshaping our environments to achieve our goals," *Philosophy & Technology*, 25/4: 589–604.

Levy, N. (2014a). *Consciousness and Moral Responsibility*. Oxford: Oxford University Press.

Levy, N. (2014b). "Consciousness, implicit attitudes and moral responsibility," *Noûs*, 48/1: 21–40.

Levy, N. (2015). "Neither fish nor fowl: Implicit attitudes as patchy endorsements," *Noûs*, 49/4: 800–23.

Levy, N. (2016). "Implicit bias and moral responsibility: Probing the data," *Philosophy and Phenomenological Research*, 93/3: 3–26.

Levy, N. (2017). "Am I a racist? Implicit bias and the ascription of racism," *Philosophical Quarterly*, 67/268: 534–51.

Levy, N. (2018). "You meta believe it," *European Journal of Philosophy*, 26/2: 814–26.

Levy, N. (2019a). "Due deference to denialism: Explaining ordinary people's rejection of established scientific findings," *Synthese*, 196/1: 313–27.

Levy, N. (2019b). "Nudge, nudge, wink, wink: Nudging is giving reasons," *Ergo: An Open Access Journal of Philosophy*, 6(10). DOI: 10.3998/ergo.12405314.0006.010

Levy, N. (2019c). "No platforming and higher-order evidence, or anti-anti-no-platforming," *Journal of the American Philosophical Association*, 5/4: 487–502.

Levy, N., & Alfano, M. (2019). "Knowledge from vice: Deeply social epistemology," *Mind*, 515: 887–915.

Lewandowsky, S., Ecker, U. K. H., Seifer, C. M., Schwarz, N., & Cook, J. (2012). "Misinformation and its correction: Continued influence and successful debiasing," *Psychological Science in the Public Interest*, 13/3: 106–31.

Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). "The role of conspiracist ideation and worldviews in predicting rejection of science," *PLOS ONE*, 8/10: e75637.

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). "The pivotal role of perceived scientific consensus in acceptance of science," *Nature Climate Change*, 3/4: 399–404.

Lewis, C. (2002). *The Dating Game: One Man's Search for the Age of the Earth*. Cambridge: Cambridge University Press.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Likens, G. E., & Bormann, F. H. (1974). "Acid rain: A serious regional environmental problem," *Science*, 184/4142: 1176–9.

Lindsay, D. S. (2015). "Replication in psychological science," *Psychological Science*, 26/12: 1827–32.

Loftus, E. F. (2003). "Make-believe memories," *The American Psychologist*, 58/11: 867–73.

Lopez, G. (2017). "Lindsey Graham, 2017: I'm tired of media portraying Trump as a kook. Graham, 2016: Trump is a kook." *Vox*, November 30. Retrieved December 17, 2019, from <https://www.vox.com/policy-and-politics/2017/11/30/16720814/lindsey-graham-trump-kook>

Lord, C. G., & Taylor, C. A. (2009). "Biased assimilation: Effects of assumptions and expectations on the interpretation of new evidence," *Social and Personality Psychology Compass*, 3/5: 827–41.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology*, 37/11: 2098–109.

Luhrmann, T. M. (2012). *When God Talks Back: Understanding the American Evangelical Relationship with God*, Reprint edition. New York: Vintage.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). "The hidden structure of overimitation," *Proceedings of the National Academy of Sciences of the United States of America*, 104/50: 19751–6.

Mackay, C. (2011). *Extraordinary Popular Delusions and the Madness of Crowds*. Petersfield, UK: Harriman House Ltd.

Malhotra, D. (2010). "(When) are religious people nicer? Religious salience and the 'Sunday Effect' on pro-social behavior," *Judgment and Decision Making*, 5/2: 138–43.

Mallapaty, S. (2018). "Paper authorship goes hyper" *Nature* Index, January 30. Retrieved September 28, 2019, from <https://www.natureindex.com/news-blog/paper-authorship-goes-hyper>

Mandelbaum, E. (2016). "Attitude, inference, association: On the propositional structure of implicit bias," *Noûs*, 50/3: 629–58.

Maoz, I., Ward, A., Katz, M., & Ross, L. (2002). "Reactive devaluation of an 'Israeli' vs. 'Palestinian' peace proposal," *The Journal of Conflict Resolution*, 46/4: 515–46.

Martin, J. L., & Desmond, M. (2010). "Political position and social knowledge," *Sociological Forum*, 25/1: 1–26.

Maruthappu, M., Watkins, J., Noor, A. M., Williams, C., Ali, R., Sullivan, R., Zeltner, T., et al. (2016). "Economic downturns, universal health coverage, and cancer mortality in high-income and middle-income countries, 1990–2010: a longitudinal analysis," *The Lancet*, 388/10045: 684–95.

Marx, K., & Engels, F. (2012). *The Communist Manifesto* (J. C. Isaac, ed.). New Haven, CT: Yale University Press.

Mascaro, O., & Sperber, D. (2009). "The moral, epistemic, and mindreading components of children's vigilance towards deception," *Cognition*, 112/3: 367–80.

Masicampo, E. J., & Lalande, D. R. (2012). "A peculiar prevalence of p values just below .05," *The Quarterly Journal of Experimental Psychology.*, 65/11: 2271–9.

McCook, A. A. (2017). "One way to boost your uni's ranking: Ask faculty to cite each other." *Retraction Watch*, August 22. Retrieved January 26, 2020, from <https://retractionwatch.com/2017/08/22/one-way-boost-unis-ranking-ask-faculty-cite/>

McCright, A. M., & Dunlap, R. E. (2011). "The politicization of climate change and polarization in the American public's views of global warming, 2001–2010," *The Sociological Quarterly*, 52/2: 155–94.

McCright, A. M., Marquart-Pyatt, S. T., Shwom, R. L., Brechin, S. R., & Allen, S. (2016). "Ideology, capitalism, and climate: Explaining public views about climate change in the United States," *Energy Research & Social Science*, 21: 180–9.

McCright, A. M., Xiao, C., & Dunlap, R. E. (2014). "Political polarization on support for government spending on environmental protection in the USA, 1974–2012," *Social Science Research*, 48: 251–60.

McGrath, S. (2008). "Moral disagreement and moral expertise." In Shafer-Landau R. (ed.), *Oxford Studies in Metaethics*, Vol. 3, pp. 87–108. New York: Oxford University Press.

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). "Recommendations implicit in policy defaults," *Psychological Science*, 17/5: 414–20.

Meltzoff, A. N. (1988). "Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli," *Developmental Psychology*, 24/4: 470–6.

Mercier, H. (2017). "How gullible are we? A review of the evidence from psychology and social science," *Review of General Psychology*, 21/2: 103–122.

Mercier, H. (2020). *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton, NJ: Princeton University Press.

Mercier, H., & Sperber, D. (2011). "Why do humans reason? Arguments for an argumentative theory," *Behavioral and Brain Sciences*, 34/02: 57–74.

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason: A New Theory of Human Understanding*. London etc.: Allen Lane.

Mercier, H., Trouche, E., Yama, H., Heintz, C., & Girotto, V. (2015). "Experts and laymen grossly underestimate the benefits of argumentation for reasoning," *Thinking & Reasoning*, 21/3: 341–55.

Mercier, H., Boudry, M., Paglieri, F., & Trouche, E. (2017). "Natural-born arguers: Teaching how to make the best of our reasoning abilities," *Educational Psychologist*, 52/1: 1–16.

Merkley, E., & Stecula, D. (2018). "Party elites or manufactured doubt? The informational context of climate change polarization," *Science Communication*, 40/2: 258–74.

Meyer, M., & Alfano, M. (forthcoming). "Fake news, conspiracy theorizing, and intellectual vice." In Alfano, M., Klein, C., & de Ridder, J. (eds.), *Social Virtue Epistemology*. Abingdon and New York: Routledge.

Miller, D. (2019). "The mystery of evangelical Trump support?," *Constellations*, 26/1: 43–58.

Miller, D. T., & McFarland, C. (1991). "When social comparison goes awry: The case of pluralistic ignorance." *Social Comparison: Contemporary Theory and Research*, pp. 287–313. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Millgram, E. (2015). *The Great Endarkenment: Philosophy for an Age of Hyperspecialization*. Oxford and New York: Oxford University Press.

Mills, C. M., & Keil, F. C. (2004). "Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth," *Journal of Experimental Child Psychology*, 87/1: 1–32.

Moles, A. (2007). "Autonomy, free speech and automatic behaviour," *Res Publica*, 13/1: 53–75.

Monroe, A. E., Brady, G. L., & Malle, B. F. (2017). "This isn't the free will worth looking for: General free will beliefs do not influence moral judgments, agent-specific choice ascriptions do," *Social Psychological and Personality Science*, 8/2: 191–9.

Moore, M. T., & Fresco, D. M. (2012). "Depressive realism: A meta-analytic review," *Clinical Psychology Review*, 32/6: 496–509.

Morin, O. (2016). *How Traditions Live and Die*. Oxford and New York: Oxford University Press.

Motta, M., Chapman, D., Stecula, D., & Haglin, K. (2019). "An experimental examination of measurement disparities in public climate change beliefs," *Climatic Change*, 154/1: 37–47.

Murphy, K. (2020). "Two moderate Liberal MPs say government should not back new coal plants." *The Guardian*, February 10.

Nadelhoffer, T., Shepard, J., Crone, D. L., Everett, J. A. C., Earp, B. D., & Levy, N. (2020). "Does encouraging a belief in determinism increase cheating? Reconsidering the value of believing in free will," *Cognition*, 203: 104342. DOI: 10.1016/j.cognition.2020.104342

Nagell, K., Olguin, R. S., & Tomasello, M. (1993). "Processes of social learning in the tool use of chimpanzees (Pan troglodytes) and human children (Homo sapiens)," *Journal of Comparative Psychology*, 107/2: 174–86.

Newport, F. (2014). "In U.S., 42% believe creationist view of human origins." *Gallup.com*. Retrieved November 24, 2019, from <https://news.gallup.com/poll/170822/believe-creationist-view-human-origins.aspx>

Nurse, M. S., & Grant, W. J. (2020). "I'll see it when I believe it: Motivated numeracy in perceptions of climate change risk," *Environmental Communication*, 14/2: 184–201.

O'Connor, C., & Weatherall, J. O. (2019). *The Misinformation Age: How False Beliefs Spread*, 1st edition. New Haven, CT: Yale University Press.

Open Science Collaboration. (2015). "Estimating the reproducibility of psychological science," *Science*, 349/6251: aac4716.

Oreskes, N., & Conway, E. M. (2011). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. London: A&C Black.

Palus, S. (2015). "Is double-blind review better?," *APS News*, 24/7.

Pasek, J., Schneider, D., Krosnick, J. A., Tahk, A., Ophir, E., & Milligan, C. (2014). "Prevalence and moderators of the candidate name-order effect evidence from statewide general elections in California," *Public Opinion Quarterly*, 78/2: 416–39.

Peck, T. (2020). "The government claims it's been 'led by the science'. So why has it led to a place far worse than elsewhere?" *The Independent*, April 13.

Pennycook, G., & Rand, D. G. (2019). "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning," *Cognition*, *The Cognitive Science of Political Thought*, 188: 39–50.

Perreault, C. (2012). "The pace of cultural evolution," *PLOS ONE*, 7/9: e45150. DOI: 10.1371/journal.pone.0045150

Pew Research Center. (2017). "Sharp partisan divisions in views of national institutions." *Pew Research Center for the People and the Press*. Retrieved December 29, 2019, from <https://www.people-press.org/2017/07/10/sharp-partisan-divisions-in-views-of-national-institutions/>

Pew Research Center. (2019). "Changing attitudes on same-sex marriage." *Pew Research Center's Religion & Public Life Project*. Retrieved December 21, 2019, from <https://www.pewforum.org/fact-sheet/changing-attitudes-on-gay-marriage/>

Pigden, C. (2007). "Conspiracy theories and the conventional wisdom," *Episteme* 4/2: 219–32.

Pomerantsev, P. (2014). "Russian and the menace of unreality." *The Atlantic*, September 10.

Prior, M., Sood, G., & Khanna, K. (2015). "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions," *Quarterly Journal of Political Science*, 10/4: 489–518.

Putnam, H. (1975). "The meaning of 'meaning'," *Minnesota Studies in the Philosophy of Science*, 7: 131–93.

Quinn, B., & Campbell, D. (2021). "Hospital incursions by Covid deniers putting lives at risk, say health leaders." *The Guardian,* January 28.

Quintana, I. O., Klein, C., Cheong, M., Sullivan, E., Reimann, R., & Alfano, M. (n.d.). *The Evolution of Vaccine Discourse on Twitter during the First Six Months of COVID-19*.

Rabb, N., Fernbach, P. M., & Sloman, S. A. (2019). "Individual representation in a community of knowledge," *Trends in Cognitive Sciences*, 23/10: 891–902.

Raphelson, S. (2018). "Survivors of mass shootings face renewed trauma from conspiracy theorists." *NPR.org*, March 28. Retrieved July 3, 2019, from <https://www.npr.org/2018/03/20/595213740/survivors-of-mass-shootings-face-renewed-trauma-from-conspiracy-theorists>

Rayner, K. (1998). "Eye movements in reading and information processing: 20 years of research," *Psychological Bulletin*, 124/3: 372–422.

Readfearn, G. (2015). "Doubt over climate science is a product with an industry behind it." *The Guardian*, March 5.

Reeves, A., McKee, M., & Stuckler, D. (2014). "Economic suicides in the Great Recession in Europe and North America," *The British Journal of Psychiatry*, 205/3: 246–7.

Richardson, V. (2017). "Climate change whistleblower alleges NOAA manipulated data to hide global warming 'pause'." *The Washington Times*, 5 February.

Richerson, P. J., & Boyd, R. (2008). *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Rieznik, A., Moscovich, L., Frieiro, A., Figini, J., Catalano, R., Garrido, J. M., Heduan, F. Á., et al. (2017). "A massive experiment on choice blindness in political decisions: Confidence, confabulation, and unconscious detection of self-deception," *PLOS ONE*, 12/2: e0171108. DOI: 10.1371/journal.pone.0171108

Roberts, R. C., & Wood, W. J. (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.

Robles, F. (2016). "Florida woman is charged with threatening Sandy Hook parent." *The New York Times*, December 7.

Rose, D., Buckwalter, W., & Turri, J. (2014). "When words speak louder than actions: Delusion, belief, and the power of assertion," *Australasian Journal of Philosophy*, 4: 1–18.

Rozenblit, L., & Keil, F. (2002). "The misunderstood limits of folk science: An illusion of explanatory depth," *Cognitive Science*, 26/5: 521–62.

Rozin, P., Scott, S., Dingley, M., Urbanek, J. K., Jiang, H., & Kaltenbach, M. (2011). "Nudge to obesity I: Minor changes in accessibility decrease food intake," *Judgment and Decision Making*, 6/4: 10.

Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.

Rust, J., & Schwitzgebel, E. (2009). "Ethicists" and nonethicists' responsiveness to student e-mails: Relationships among expressed normative attitude, self-described behavior, and empirically observed behavior,' *Philosophical Psychology*, 22/6: 711–25.

Saghai, Y. (2013). "Salvaging the concept of nudge," *Journal of Medical Ethics*, 39/8: 487–93.

Schacter, D. L. (1996). *Searching For Memory: The Brain, the Mind, and the Past.* New York: Basic Books.

Schaffner, B. F., & Luks, S. (2018). "Misinformation or expressive responding? What an inauguration crowd can tell us about the source of political misinformation in surveys," *Political Opinion Quarterly*, 82/1: 135–47.

Schaik, C., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C., Singleton, I., Suzuki, A., et al. (2003). "Orangutan cultures and the evolution of material culture," *Science (New York, N.Y.)*, 299: 102–5.

Schechtman, M. (1996). *The Constitution of Selves.* New York: Cornell University Press.

Schliesser, E. (2019). "Synthetic philosophy," *Biology and Philosophy*, 34/2: 19.

Schliesser, E., & Winsberg, E. (2020). "Climate and coronavirus: The science is not the same." *New Statesman*, March 23.

Schmidt, A. T. (2019). "Getting real on rationality—behavioral science, nudging, and public policy," *Ethics*, 129/4: 511–43.

Scholl, B. (2017). "Reliability in psychology: Means versus ends," *APS Observer*, 30: 38–9.

Schuldt, J. P., Konrath, S. H., & Schwarz, N. (2011). "'Global warming' or 'climate change'? Whether the planet is warming depends on question wording," *Public Opinion Quarterly*, 75/1: 115–24.

Schwitzgebel, E. (2009a). "Do ethicists steal more books?," *Philosophical Psychology*, 22/6: 711–25.

Schwitzgebel, E. (2009b). "Knowing your own beliefs," *Canadian Journal of Philosophy*, 39/sup1: 41–62.

Schwitzgebel, E. (2010). "Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief," *Pacific Philosophical Quarterly*, 91/4: 531–53.

Sellars, W. S. (1962). "Philosophy and the scientific image of man." In Colodny, R. (ed.), *Science, Perception, and Reality*, pp. 35–78. New York: Humanities Press/ Ridgeview.

Shafir, E., & LeBoeuf, R. A. (2002). "Rationality," *Annual Review of Psychology*, 53/1: 491–517.

Shea, N. (2009). "Imitation as an inheritance system," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364/1528: 2429–43.

Sher, S., & McKenzie, C. R. M. (2006). "Information leakage from logically equivalent frames," *Cognition*, 101/3: 467–94.

Shtulman, A. (2006). "Qualitative differences between naïve and scientific theories of evolution," *Cognitive Psychology*, 52/2: 170–94.

Simons, D. J., & Chabris, C. F. (1999). "Gorillas in our midst: sustained inattentional blindness for dynamic events," *Perception*, 28/9: 1059–74.

Simons, Daniel J., & Levin, D. T. (1997). "Change blindness," *Trends in Cognitive Sciences*, 1/7: 261–7.

Simons, Daniel J., & Levin, D. T. (1998). "Failure to detect changes to people during a real-world interaction," *Psychonomic Bulletin & Review*, 5/4: 644–9.

Smith, N. C., Goldstein, D. G., & Johnson, E. J. (2013). "Choice without awareness: Ethical and policy implications of defaults," *Journal of Public Policy & Marketing*, 32/2: 159–72.

Somin, I. (2013). *Democracy and Political Ignorance: Why Smaller Government Is Smarter*. Stanford: Stanford University Press.

Sosa, E. (1991). *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.

Sperber, D. (1996). *Explaining Culture*, 1st edition. Oxford, UK and Cambridge, MA: Blackwell Publishers.

Sperber, D. (1997). "Intuitive and reflective beliefs," *Mind & Language*, 12/1: 67–83.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). "Epistemic vigilance," *Mind & Language*, 25/4: 359–93.

Stanovich, K. E. (2004). *The Robot's Rebellion*. Chicago: University of Chicago Press.

Stanovich, K. E. (2018). "Miserliness in human cognition: The interaction of detection, override and mindware," *Thinking & Reasoning*, 24/4: 423–44.

Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, Mass.: MIT Press.

Sterelny, K. (2015). "Content, control and display: The natural origins of content," *Philosophia*, 43/3: 549–64.

Sterelny, K. (2017). "Cultural evolution in California and Paris," *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 62: 42–50.

Sterelny, K. (2018). "Religion re-explained," *Religion, Brain & Behavior*, 8/4: 406–25.

Stich, S. P. (1978). "Beliefs and subdoxastic states," *Philosophy of Science*, 45/December: 499–518.

Straumsheim, C. (2017). "Librarian's list of 'predatory' journals reportedly removed due to 'threats and politics'." *Inside Higher Education*, January 18.

Strevens, M. (2020). *The Knowledge Machine: How Irrationality Created Modern Science*. New York: Liveright.

Suls, R. (2017). "More Democrats than Republicans call Russia 'greatest danger' to US." *Pew Research Center*. Retrieved December 22, 2019, from <https://www.pewresearch.org/fact-tank/2017/04/20/share-of-democrats-calling-russia-greatest-danger-to-u-s-at-its-highest-since-end-of-cold-war/>

Sunstein, C. R. (2006). *Infotopia: How Many Minds Produce Knowledge*. Oxford ; New York: Oxford University Press.

Sunstein, C. R., & Zeckhauser, R. (2011). "Overreaction to fearsome risks," *Environmental and Resource Economics*, 48/3: 435–49.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*. New York: Doubleday.

Taber, C. S., & Lodge, M. (2006). "Motivated skepticism in the evaluation of political beliefs," *American Journal of Political Science*, 50/3: 755–69.

Talmont-Kaminski, K. (2020). "Epistemic vigilance and the science/religion distinction," *Journal of Cognition and Culture*, 20/1–2: 88–99.

Tappin, B. M., Pennycook, G., & Rand, D. (2020). "Rethinking the link between cognitive sophistication and identity-protective bias in political belief formation." *Journal of Experimental Psychology: General.* doi: 10.1037/xge0000974

Tennie, C., Call, J., and Tomasello, M. (2009). "Ratcheting up the ratchet: On the evolution of cumulative culture," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364/1528: 2405–15.

Tetlock, P. E. (2003). "Thinking the unthinkable: sacred values and taboo cognitions," *Trends in Cognitive Sciences*, 7/7: 320–4.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness.* New Haven, CT: Yale University Press.

Todd, P. M., & Gigerenzer, G. (2012). *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press.

Toscano, N. (2018). "'Rallying cry': Mining giant BHP renews calls for carbon price." *The Sydney Morning Herald*, October 22. Retrieved December 27, 2019, from <https://www.smh.com.au/business/companies/rallying-cry-mining-giant-bhp-renews-calls-for-carbon-price-20181022-p50b5e.html>

Tversky, A., & Kahneman, D. (1981). "The framing of decisions and the psychology of choice," *Science*, 211/4481: 453–8.

Uhlmann, E. L., & Cohen, G. L. (2005). "Constructed criteria: Redefining merit to justify discrimination," *Psychological Science*, 16/6: 474–80.

Uscinski, J. E., & Parent, J. M. (2014). *American Conspiracy Theories.* Oxford: Oxford University Press.

Vaidis, D., & Sleegers, W. (2018). "Large scale registered replication project – cognitive dissonance: Induced compliance paradigm with counterattitudinal essay." *Open Science Foundation*. Retrieved December 20, 2019, from <https://osf.io/9xsmj/>

van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2015). "The scientific consensus on climate change as a gateway belief: Experimental evidence," *PLOS ONE*, 10/2: e0118489. DOI: 10.1371/journal.pone.0118489

van der Linden, S. L., Leiserowitz, A., & Maibach, E. (2019). "The gateway belief model: A large-scale replication," *Journal of Environmental Psychology*, 62: 49–58.

van der Linden, S. L., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). "Inoculating the public against misinformation about climate change," *Global Challenges*, 1/2: 1600008.

van der Linden, S. L., Panagopoulos, C., Azevedo, F., & Jost, J. T. (2021). "The paranoid style in American politics revisited: An ideological asymmetry in conspiratorial thinking," *Political Psychology*, 42/1.

Van Leeuwen, N. (2014). "Religious credence is not factual belief," *Cognition*, 133/3: 698–715.

Van Leeuwen, N. (2018). "The factual belief fallacy," *Contemporary Pragmatism*, 15/3: 319–43.

Van Leeuwen, N. (forthcoming). *Imagination, Belief, and Religious Credence*, Harvard University Press.

Vohs, K. D., & Schooler, J. W. (2008). "The value of believing in free will: encouraging a belief in determinism increases cheating," *Psychological Science*, 19/1: 49–54.

Vosoughi, S., Roy, D., & Aral, S. (2018). "The spread of true and false news online," *Science*, 359/6380: 1146–51.

Walker, R., & Rocha da Silva, P. (2015). "Emerging trends in peer review—a survey," *Frontiers in Neuroscience*, 9. Frontiers. DOI: 10.3389/fnins.2015.00169

Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). "Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus," *Journal of Personality and Social Psychology*, 92/5: 821–33.

Whalen, A. (2019). "Are aliens real? One-third of Americans think alien UFOs have visited earth." *Newsweek*, September 6.

WHO. (2019). "Ten threats to global health in 2019." Retrieved December 6, 2019, from <https://www.who.int/emergencies/ten-threats-to-global-health-in-2019>

Wilkinson, T. M. (2013). "Thinking harder about Nudges," *Journal of Medical Ethics*, 39/8: 486–486.

Williams, D. (2021). "Socially adaptive belief," *Mind & Language* 36: 333–54.

Winsberg, E. (2018). *Philosophy and Climate Science*. Cambridge, United Kingdom ; New York, NY: Cambridge University Press.

Winsberg, E., Huebner, B., & Kukla, R. (2014). "Accountability and values in radically collaborative research," *Studies in History and Philosophy of Science Part A*, 46: 16–23.

Wolf, N. (2019). *Outrages: Sex, Censorship and the Criminalisation of Love*. London: Virago Press Ltd.

Wolterstorff, N. (1996). *John Locke and the Ethics of Belief*. Cambridge: Cambridge University Press.

Wulf, K. (2019). "What Naomi Wolf and Cokie Roberts teach us about the need for historians." *Washington Post*, June 6.

Xygalatas, D. (2012). "Effects of religious setting on cooperative behavior: A case study from Mauritius," *Religion, Brain and Behavior*, 3: 91–102.

Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

Zaller, J., & Feldman, S. (1992). "A simple theory of the survey response: Answering questions versus revealing preferences," *American Journal of Political Science*, 36/3: 579.

Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zollman, K. (2010). "The epistemic benefit of transient diversity," *Erkenntnis*, 72/1: 17–35.

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.