IV

What Reason Can and Cannot Do

In Chapters 7 through 10, we have developed a novel interactionist approach to reason. According to this approach, the function of reason is to produce and evaluate justifications and arguments in dialogue with others. For the standard intellectualist approach, the function of reason is to reach better beliefs and make better decisions on one's own. As we'll see, the two approaches yield sharply contrasting predictions that we are now in a position to test. Is reason objective or biased? Is it demanding or lazy? Does it help the lone reasoner? Does it yield better or worse results in interactive contexts? Looking at the evidence with these questions in mind throws new light on the promises and dangers of reason.

2017. Harvard University Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

EBSCO Publishing: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY 1491806; Hugo Mercier, Dan Sperber; The Enigma of Reason Account:ehost.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

11

Why Is Reasoning Biased?

There is much misunderstanding about the way to test adaptive hypotheses. It might seem that in order to understand what reason is for, why it evolved, we must be able to find out how our ancestors reasoned, using archeological or genetic data. Fortunately, million-year-old skulls and gene sequencing are not indispensable to test hypotheses about the function of an evolved mechanism. Think about it: we can tell what human eyes are for without knowing anything about when and how they evolved. What matters is the existence of a match between the function of an organ, or a cognitive mechanism, and its structure and effects. Do the features of the eye serve its function well? By and large, yes. Do eyes achieve their function well? By and large, yes.

We can use the same logic to guide our examination of the data on human reason. Do the features of human reason serve best the functions posited by the intellectualist or the interactionist approach? Which functions does it achieve best? In many cases reason couldn't serve both functions well at the same time, so there will be plenty of evidence to help us decide between the two approaches. We'll start our tour of what reason can and cannot do with a historical case. How does a certified scientific genius reason? Does reason help him discard misguided beliefs and reach sounder conclusions?

When Linus Pauling received the American Chemical Society's Award in Pure Chemistry—he was only thirty years old—a senior colleague predicted that he would win a Nobel Prize.[1] Actually, Pauling won *two* Nobel Prizes—Chemistry and Peace—joining Marie Curie in this exclusive club. As a serious contender in the race to discover the structure of DNA, he narrowly missed a third Nobel Prize. Indeed, James Watson was long afraid to be beaten

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

by Pauling's "prodigious mind."[2] A former student who would also become a Nobel laureate described him as a "god-like, superhuman, great figure."[3] Undoubtedly, Linus Pauling had a great mind, and he was steeped in the most rigorous scientific tradition.

When he was not busy winning Nobel Prizes, Pauling sometimes pondered the powers of vitamin C. At first, he advocated heavy daily doses as a prophylactic against colds and other minor ailments, stopping short of recommending vitamin C for the treatment of serious diseases. This changed when he met Ewan Cameron, a surgeon who had conducted a small study demonstrating the positive effects of vitamin C on cancer patients—or so it seemed. Pauling and Cameron joined forces to defend the potential of vitamin C as a treatment for cancer. Thanks to Pauling's clout and to his continuous efforts, the prestigious Mayo Clinic agreed to conduct a large-scale, tightly controlled trial.[4] Unfortunately for Pauling and Cameron—and for the cancer patients—the results were negative. Vitamin C had no effect whatsoever.[5]

At this point, Pauling could have objectively reviewed the available evidence: on the one hand a fringe theory and a small, poorly controlled study; on the other hand, the medical consensus and a large, well-controlled clinical trial. On the basis of this evidence, the vast majority of researchers and doctors concluded that vitamin C had no proven effects on cancer. But Pauling did not reason objectively. He built a partisan case. The first Mayo Clinic study was dismissed because its participants had, according to Pauling, not been selected properly. When a second study was performed that addressed this issue,[6] Pauling made up another problem: the new patients had not received vitamin C for an "indefinite time."[7] Pauling's requirements did not match any standard cancer research, only fitting the small trial Cameron had performed.

The most egregious example of biased reasoning on Pauling's part is found in an article, published a few years later, in which he advanced three criteria to evaluate clinical trials of cancer treatments.[8] While "most of the reported results of clinical trials of cohorts of cancer patients satisfy these criteria of validity,"[9] there was one black sheep, "a reported clinical trial that fails on each of the three criteria for validity."[10] Can you guess what this outlier was? A study "*described* as a randomized double-blind comparison of vitamin C (10 g per day) and a lactose placebo" (emphasis added).[11] The study singled out as the only flawed cancer study in a sample of several hundred is the

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Mayo Clinic study that embarrassed Pauling in front of the whole scientific community.

There is no reason to accuse Pauling of conscious intellectual dishonesty. He took high doses of vitamin C daily, and his wife used the same regimen to fight—unsuccessfully—her stomach cancer. Still, for most observers, Pauling's evaluation of the therapeutic efficacy of vitamin C is a display of selective picking of evidence and partisan arguments. Even when he was diagnosed with cancer despite having taken high doses of vitamin C for many years, he did not admit defeat, claiming that the disease would have struck earlier without it.[12]

Pauling may have erred further than most respected scientists in his unorthodox beliefs, but his way of reasoning is hardly exceptional—as anyone who knows scientists can testify, they are not paragons of objectivity (more on this in Chapter 18). Undoubtedly, even the greatest minds can reason in the most biased way.

Is Bias Always Bad?

How should cognitive mechanisms in general go about producing sound beliefs? Part of the answer, it seems, is that they should be free of bias. Biases have a bad press, in part because a common definition is "inclination or prejudice for or against one person or group, especially in a way considered to be unfair."[13] However, psychologists often use the term in a different manner, to mean a systematic tendency to commit some specific kind of mistakes. These mistakes need not have any moral, social, or political overtones.

A first kind of bias simply stems from the processing costs of cognition. Costs can be reduced by using heuristics, that is, cognitive shortcuts that are generally reliable but that in some cases lead to error. A good example of this is the "availability heuristic" studied by Tversky and Kahneman.[14] It consists in using the ease with which an event comes to mind to guess its actual frequency. For instance, in one experiment participants were asked whether the letter *R* occurs more frequently in first or third position in English words. Most people answered that *R* occurs more often in first position, when actually it occurs more often in third position. The heuristic the participants used was to try to recall words beginning with *R* (like "river") and words with *R* in

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

the third place (like "bored") and assume that the ease with which the two kinds of words came to mind reflected their actual frequency. In the particular case of the letter *R* (and of seven other consonants of the English alphabet), the availability heuristic happens to be misleading. In the case of the other thirteen consonants of the English alphabet, the same heuristic would give the right answer.

Although the availability heuristic can be described as biased—it does lead to systematic errors—its usefulness is clear when one considers the alternative: trying to count *all* the words one knows that have *R* as the first or third letter. While the heuristic can be made to look bad, we would be much more worried about a participant who would engage in this painstaking process just to answer a psychology quiz. Moreover, the psychologist Gerd Gigerenzer and his colleagues have shown that in many cases using such heuristics not only is less effortful but also gives better results than using more complex strategies.[15] Heuristics, Gigerenzer argues, are not "quick-and-dirty"; they are "fast-and-frugal" ways of thinking that are remarkably reliable.

The second type of bias arises because not all errors are created equal.[16] More effort should be made to avoid severely detrimental errors—and less effort to avoid relatively innocuous mistakes.

Here is a simple example illustrating how an imbalance in the cost of mistakes can give rise to adaptive biases. Bumblebees have cognitive mechanisms aimed at avoiding predators. Among their predators are crab spiders, small arachnids that catch the bees when they forage for nectar. Some crab spiders camouflage themselves by adopting the color of the flowers they rest on: they are cryptic. To learn more about the way bumblebees avoid cryptic predators, Thomas Ings and Lars Chittka created little robot spiders.[17] All the robots rested on yellow flowers, but some of them were white (noncryptic) while others were yellow (cryptic). To simulate the predation risk, Ings and Chittka built little pincers that held the bees captive for two seconds when they landed on a flower with a "spider."

In the first phase of the experiments, two groups of bumblebees, one facing cryptic spiders and the other facing noncryptic spiders, had multiple opportunities to visit the flowers and to learn which kind of predators they were dealing with. Surprisingly, both groups of bumblebees very quickly learned to avoid the flowers with the spiders—even when the spiders were cryptic.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Yet the camouflage wasn't ineffective: to achieve the same ability to detect spiders, the bumblebees facing camouflaged spiders spent nearly twice as long inspecting each flower. This illustrates the cost of forming accurate representations of one's environment: the time spent inspecting could not be spent harvesting.

But there is also an asymmetry in the costs of mistakenly landing on a flower with a spider (high cost) versus needlessly avoiding a spider-free flower (low cost). This second asymmetry also affected the bumblebees' behavior. On the second day of the experiment, the bees had learned about the predation risks in their environment. Instead of spending forever (well, 0.85 seconds) inspecting every flower to make sure it carried no spider, the bumblebees facing the cryptic spiders settled for a higher rate of "false alarms": they were more likely than the other bees to avoid flowers on which, in fact, there were no spiders.

This experiment illustrates the exquisite ways in which even relatively simple cognitive systems adjust the time and energy they spend on a cognitive task (ascertaining the presence of a spider on a flower) to the difficulty of the task on the one hand, and to the relative cost of a false negative (assuming there is no spider when there is) and of a false positive (assuming there is a spider when there is not) on the other hand. This difference in cost results in a bias: making more false positive than false negative errors. This bias, however, is beneficial.

From Surprise to Falsification

While the example of the bumblebees illustrates that bias may be beneficial, it speaks of very specific costs: the costs of predation versus the costs of missing a feeding opportunity. Are there cognitive biases that would be useful more generally?

A main goal of cognitive mechanisms is to maintain an accurate representation of the organism's environment, or at least of relevant aspects of it. It could be argued that only the future, more particularly the immediate future, of the environment is directly relevant: it determines both what may happen to the organism and what the organism can do to modify the environment in its favor. The past and present of the environment may be

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

relevant too, but only indirectly, by providing the only evidence available regarding the environment's future. Another way of making the same point is to say that a main goal of cognition is to provide the organism with accurate expectations regarding what may happen next. Cognitive mechanisms should pay extra attention to any information that goes against their current expectations and use this information to revise them appropriately. Information that violates expectations causes, at least in humans, a typical reaction: surprise. The experience of surprise corresponds to the sudden mobilization of cognitive resources to readjust expectations that have just been challenged.

Indeed, paying due attention to the unexpected is so ingrained that we are surprised by the lack of surprise in others. If your friend Joe, upon encountering two cats that are similar except that one meows and the other talks, fails to pay more attention to the talking cat, you'll suspect there's something wrong with him. Even one-year-old babies expect others to share their surprise. When they see something surprising, they point toward it to share their surprise with nearby adults. And they keep pointing until they obtain the proper reaction or are discouraged by the adults' lack of reactivity.[18]

This doesn't mean that people seek out surprises. If you are lucky enough to find yourself in a nice environment where you know what to expect, deviations from these expectations—surprises—are likely to be bad news (hence the fake Chinese curse, "May you live in interesting times"). However, even if you don't like surprises, when the environment has something surprising in stock for you, you're better off learning about it as early as possible. Being fired is a very bad surprise; better be surprised by indications that you might be fired before it actually happens. As a result, people favor sources of surprising information over sources of repetitive information (provided they are equally reliable): they buy the newspaper with the proverbial "Man Bites Dog" headline, not "Dog Bites Man."

If reason is an instrument of individual cognition aimed at better beliefs and decisions, it should be biased toward information that violates our expectations. Reason should look for counterarguments to our generalizations, reasons to question our decisions, memories that clash with our current beliefs. Failing to pay attention to negative evidence and arguments is generally more costly than failing to pay attention to their positive counterparts.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

This insight is so essential that it is one of the few themes that we find again and again in scholars' admonitions about the best way to reach sound beliefs. In the twelfth century, Robert Grosseteste suggested that to eliminate faulty hypotheses, one should take a hypothesis and see if anything follows from it that is illogical or contrary to fact.[19] Four centuries later, Francis Bacon overturned the scholastic tradition to which Grosseteste belonged by putting observation at the center of his philosophy. Still, we find again the idea of looking for counterexamples. Observation should not be the "childish" enumeration of cases that fit with the researcher's hypothesis. Scholars are to look specifically for instances that can prove their hypothesis wrong: one must "analyse nature by proper rejection and exclusion."[20]

Another four centuries later, it is Karl Popper's turn to challenge dominant views of the day, distant descendants of Bacon's philosophy. Popper does so by stressing more than ever before the importance of falsification. For Popper, what demarcates science from other forms of knowledge is not that scientific theories are verifiable but that they are falsifiable: in principle, a single observation of an event that, according to a given theory, could never occur would falsify the theory. In practice, of course, as Popper well recognized, a single observation is not enough to cause scientists to abandon a theory—after all, the observation itself might be mistaken. Still, when falsifying observations are numerous and reliable enough, then the theory must at least be revised if not abandoned. Scientists produce theories that are at risk of being falsified. They improve their theories by looking for falsifying evidence, by rejecting falsified theories, and by holding on only to theories that have withstood repeated attempts at falsifying them.

For all their differences, these scholars agree that counterexamples and other violations of expectations, by allowing us to discard misguided beliefs, play a crucial role in the accumulation of knowledge. As Popper put it, "In searching for the truth, it may be our best plan to start by criticizing our most cherished beliefs."[21]

The Confirmation Bias

Peter Wason devised the clever experimental tasks that were to have such an impact on the psychology of reasoning at University College in London, within

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

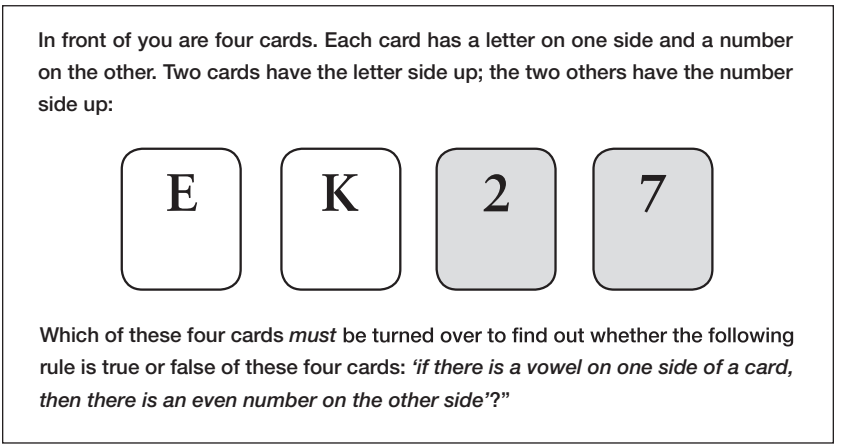

Figure 17. The standard Wason selection task.

walking distance of the London School of Economics where Popper was teaching. The proximity was more than geographic. Wason extended Popper's insight about scientific theories to everyday reasoning, asking: Do people rely on falsification to arrive at better beliefs?

The selection task, which was discussed in Chapter 2 (see Figure 17 for a reminder), was precisely meant to simulate aspects of scientific thinking. Participants, Wason thought, must select cards that provide the right kind of evidence to test the rule, just as scientist must identify the right kind of evidence to test their hypotheses. As we saw, the right selection of cards in this example is that of the card with the vowel E and that of the card with an odd number 7—these are the only two cards that might falsify the rule (if there is an odd number on the other side of the E, or a vowel on the other side of the 7). Most participants, however, fail to select the 7 card. Many select the card with the even number 2 presumably to see whether it has a vowel on the other side. But since the rule does not say that it should, this card is irrelevant to deciding whether the rule is true or false.

On the basis of this and other experiments, Wason suggested that participants suffered from a bias that would later be called "confirmation bias":[22] "The results do suggest . . . that even intelligent adults do not readily adopt a scientific attitude to a novel problem. They adhere to their own explanation with remarkable tenacity when they can produce confirming evidence for them."[23]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Actually, the experiment that convinced so many people that there is a confirmation bias is not such a straightforward example of this bias. Wason and many of his followers made a logical mistake in claiming that participants try to "verify" or "confirm" the rule instead of falsifying it: it is, in fact, exactly the same cards that can falsify the rule if it is false or verify it if it is true. In the preceding example, for instance, the E and the 7 cards could each, when turned over, reveal that the rule is false. If neither falsifies the rule, then the two same cards jointly prove that the rule is true. The quite common selection of the card with the even number 2 not only fails to falsify the rule, it also fails to confirm it. Selecting it is not a confirmation strategy at all. So people who succeed on the task do not show a greater disposition to falsify than to confirm, and people who fail do not show a greater disposition to confirm than to falsify.

If the choice of any given card cannot reveal whether people have a confirmation bias, the *way* they choose their cards can. People's selection of cards, we saw, is rapidly determined not by reason but by intuitions about the relevance of the different cards. After this initial, intuitive reaction, participants tend to reason long and hard about their choice. However, they do not reason as much about each of the four cards. By tracking participants' gaze, researchers established that participants spend all their time thinking just about the cards made relevant by the rule, the cards that they have already selected.[24] Moreover, when participants are asked to think aloud, it becomes clear that they mostly think of reasons supporting their intuitive choice.[25] Here's the real confirmation bias: instead of finding reasons for and against each card, participants find plenty of reasons supporting their initial card choice, neglecting reasons to pick other cards, or reasons to not pick the cards initially chosen.

We will soon argue that this confirmation bias is in fact best understood as a "myside bias," but let's first look at a small sample of the rich evidence demonstrating its existence.[26]

Deanna Kuhn, a pioneering scholar of argumentation and cognition, asked participants to take a stand on various social issues—unemployment, school failure, and recidivism. Once the participants had given their opinion, they were asked to justify it. Nearly all participants obliged, readily producing reasons to support their point of view. But when they were asked to produce

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

counterarguments to their own view, only 14 percent were consistently able to do so, most drawing a blank instead.[27]

Ziva Kunda led participants to believe that extroverts are more likely to be successful than introverts. On a later memory task, these participants found it much easier to recall memories of their own extroverted rather than introverted behavior. Another group of participants was led to believe that it is introverts who are more likely to be successful than extroverts. They found it easier to recall memories of introverted rather than extroverted behavior. Both groups were simply seeking reasons to believe that they had the qualities that would make them successful.[28]

Charles Taber and Milton Lodge gave participants a variety of arguments on controversial issues, such as gun control or affirmative action, and asked them to list their thoughts relative to the arguments.[29] They divided the participants into two groups: those with low and those with high knowledge of political issues. The low-knowledge group exhibited a solid confirmation bias: they listed twice as many thoughts supporting their side of the issue than thoughts going the other way. But knowledge did not protect the participants from bias. The participants in the high-knowledge group found so many thoughts supporting their favorite position that they gave *none* going the other way. Greater political knowledge only amplified their confirmation bias.

The list could go on for pages (indeed for chapters or books, even). Moreover, as the example of Pauling suggests, it is not only ordinary participants who fall prey to the confirmation bias. Being gifted, focused, motivated, or open minded is no protection against the confirmation bias.[30] A small industry of experiments has busily demonstrated the prevalence and robustness of what is "perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning."[31] As the journalist Jon Ronson quipped, "Ever since I learnt about confirmation bias I've started seeing it everywhere."[32]

A Challenge for the Intellectualist Approach

Psychologists agree that the confirmation bias is prevalent. They also agree that it is a bad thing. The confirmation bias is "irrational,"[33] and it "thwart[s]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

the ability of the individual to maximize utility."[34] It is the "bias most pivotal to ideological extremism and inter- and intragroup conflict."[35] Raymond Nickerson aptly summarizes the common view:

> Most commentators, by far, have seen the confirmation bias as a human failing, a tendency that is at once pervasive and irrational. It is not difficult to make a case for this position. The bias can contribute to delusions of many sorts, to the development and survival of superstitions, and to a variety of undesirable states of mind, including paranoia and depression. It can be exploited to great advantage by seers, soothsayers, fortune tellers, and indeed anyone with an inclination to press unsubstantiated claims. One can also imagine it playing a significant role in the perpetuation of animosities and strife between people with conflicting views of the world.[36]

A damning assessment indeed. Moreover, this bad thing, far from being hidden in the recesses of human psychology, is well in view. It doesn't take a very shrewd or cynical observer of human nature to realize that humans have a confirmation bias. Why did Bacon, for instance, take such care to warn against the dangers of the "childish" enumeration of instances? Because he was well aware of people's tendency to confirm their beliefs:

> The human understanding when it has once adopted an opinion . . . draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects; in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate.[37]

Bacon, writing more than two centuries before Darwin's discovery of natural selection, could merely observe that people have a confirmation bias and that this bias hinders the acquisition of sound beliefs. For contemporary defenders of the intellectualist approach to reason who take it for granted that reason is an outcome of Darwinian natural selection, the very existence of the confirmation bias presents a radical challenge.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

In particular, a defender of the intellectualist approach should accept the following three claims: (1) reason has a confirmation bias, (2) the confirmation bias makes it harder for reason to help the lone reasoner arrive at better beliefs and better decisions, and (3) the main function of reason is to arrive at better beliefs and better decisions. This makes as much sense as accepting the three following claims: (1) the elk's antlers are enormous, (2) the size of these antlers makes it harder for the elks to avoid predators, and (3) the function of the enormous antlers is to avoid predators.

The only ways out of this conundrum that come to mind are nonstarters. A first way out would be to argue that the confirmation bias is an unavoidable feature of reason. For instance, the extra weight that the bones, muscles, and feathers of avian wings add to the body of a bird makes it harder, not easier, for that body to fly. Yet since neither weightless wings nor wingless flight is an option, the weight of wings does not in itself raise a problem for evolutionary theory. But why would having a confirmation bias be necessary for reason to function at all?

A second escape route would be to claim that the confirmation bias serves a secondary function. Some features of an adaptation may hinder its main function but be well explained by the fact that they serve a secondary function. Razorbills, for instance, have a wing area relative to body mass ("wing loading") that is too low for optimal flight. This feature, however, is explained by the fact that their wings are also adapted for underwater propulsion.[38] Unlike razorbills' low wing loading, the confirmation bias doesn't just make reason's alleged main function a bit harder to achieve; it works directly against it. Moreover, there is no particularly plausible secondary function of reason that would explain the confirmation bias the way underwater propulsion explains low wing loading in razorbills. For the intellectualist approach, the confirmation bias should be a deep puzzle.

Can Intuitions Be Blamed for Confirmation Bias?

Supporters of the intellectualist view who are aware of how problematic the confirmation bias is still have one option: deny that the confirmation bias is a feature of reason proper, shifting the blame from reason to intuition. Indeed, when the dual process approach, briefly discussed in Chapter 2, became

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

popular, one of its main attractions was that it seemed to help solve the enigma of reason: weaknesses of reasoning could now be blamed on type 1 intuitive processes, while type 2 reasoning proper could be exonerated. Intuitions, it was claimed, make mistakes that it is one function of reason to correct.[39] Stanovich, for instance, in his 2004 book *The Robot's Rebellion,* listed the confirmation bias among other alleged cognitive weaknesses of intuition.[40] Similarly, Evans suggested that the confirmation bias resulted from a more general "bias to think about positive rather than negative information."[41]

This solution—blaming the intuitions—doesn't make much evolutionary sense. Mechanisms of intuitive inference guide our thoughts and actions; natural selection has honed some of these mechanisms for hundreds of millions of years. Our survival and reproduction very much depend on the quality of the information provided by intuitions. As we saw, some specific biases may on the whole be advantageous when they lower the costs of cognition or make less likely particularly costly kinds of mistake. The confirmation bias carries none of these advantages.

Unsurprisingly, then, no confirmation bias emerges from studies of animal behavior. A mouse bent on confirming its belief that there are no cats around and a rat focusing its attention on breadcrumbs and ignoring other foods to confirm its belief that breadcrumbs are the best food would not pass on their genes to many descendants. Foraging behaviors adapt to changing environments. Animals abandon food patches as soon as they expect to find better elsewhere.[42] Human intuitions are no worse than the inferences of other animals. Our ancestors passed on to us abilities for foraging and avoiding predators, as well as a great variety of other inferential devices that do not suffer from a confirmation bias.[43]

If anything, as we argued earlier, we should expect intuitions to be biased toward disconfirming, surprising information. An experiment conducted by psychologist Thomas Allen and his colleagues offers a nice demonstration of the respective biases of intuitions and reason.[44] The participants were shown a picture of an individual and two statements describing his behavior, and they were then asked to form an impression of this individual. Some participants saw the picture of "a young adult Black male wearing a black headband and dark sunglasses"[45] followed by two statements: "Swore at the salesgirl" and "Gave up his seat on the crowded subway to the elderly man." If you are

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

familiar with the stereotypes of young black males in the United States, you will have figured out that the first statement was designed to comport with most participants' expectations, and the second to be surprising.

In order to test the role of intuition in impression formation, half the participants were stopped from using reason—by having to hold in mind a long string of digits, a task that monopolizes resources necessary for sustained reasoning. These participants, then, were guided by their intuitions. As we would have predicted, these participants paid more attention to the surprising statement. By contrast, participants who could reason paid more attention to the unsurprising statement. Intuitions aimed at gathering the most useful information while reasoning aimed at confirming the participants' stereotypes.

The Myside Bias—and What Is It For?

So far, we have taken for granted that the bias described was a *confirmation* bias, a bias to confirm whatever view one happens to be entertaining. However, some experiments reveal clearly that this is not a good description of what reasoning does. For instance, we saw earlier that participants have trouble finding counterarguments to their favorite theories. But when participants are asked to reason about ideas they disagree with, they easily find counterarguments.[46]

What these results—and many others[47]—show is that people have no general preference for confirmation. What they find difficult is not looking for counterevidence or counterarguments in general, but only when what is being challenged is their own opinion. Reasoning does not blindly confirm any belief it bears on. Instead, reasoning systematically works to find reasons for our ideas and against ideas we oppose. It always takes our side. As a result, it is preferable to speak of a *myside bias* rather than of a confirmation bias.[48]

This being cleared up, let's recap. A lot of evidence shows that reasoning has a myside bias. Reason rarely questions reasoners' intuitions, making it very unlikely that it would correct any misguided intuitions they might have. This is pretty much the *exact opposite* of what you should expect of a mechanism that aims at improving one's beliefs through solitary ratiocination. There is no obvious way to explain the myside bias from within the intellectualist approach to reasoning.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

A strong, universal bias is unlikely to be a mere bug, a bad thing. Instead, it is more likely to be a useful feature. What goal could the myside bias serve? The analogy of the lone reasoner as scientist proposed by Wason does not fit, but another analogy does throw some light on this mysterious bias.

Cicero may be Western civilization's most illustrious lawyer, his writings on rhetoric echoing from the Roman senate down through the halls of medieval and modern universities. This is how, in *De Inventione,* he advised orators to conclude a speech:

> It will be serviceable both to run over the arguments which you yourself have employed separately, and also (which is a matter requiring still greater art) to unite the opposite arguments with your own; and to show how completely you have done away with the arguments which were brought against you. And so, by a brief comparison, the recollection of the hearer will be refreshed both as to the confirmation which you adduced, and as to the reprehension which you employed.[49]

In other words, when you want to convince someone, give only arguments that support your position or that counter the position you oppose. Cicero is bluntly advocating for the myside bias. And it makes complete sense. If a lawyer starts arguing against her client or for the other side, she'll soon be out of business.[50]

The lawyer analogy brings to mind a context in which persuasion is paramount and the myside bias makes obvious sense: when defending a point of view, the myside bias is a good thing.[51] *It is a feature, not a bug.* This fits with the prediction of the interactionist approach. If the function of reasoning, when it produces reasons, is to justify one's actions or to convince others, then it *should* have a myside bias.

Being Our Own Lawyers

If the myside bias is so ingrained in reason, why should Cicero have to remind us to have a myside bias? While the "reason is a lawyer" analogy can be illuminating, especially by contrast with the more typical analogy of reason as a scientist, it shouldn't be pushed too far. For instance, unlike typical

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

reasoners, lawyers often argue for positions they personally do not endorse. Their "side" is that of the client who employs them. Finding arguments for a position we do not support, or even one we disagree with, is difficult. It takes skills and training. We may be lawyers, but only when it comes to defending beliefs and decisions we actually endorse.

There are other reasons not to take the lawyer analogy too far. Besides reason, many other cognitive mechanisms are at play when lawyers prepare their plea. When they consciously suppress arguments against their client or use such arguments to anticipate what the other party might say, they rely on strategic planning rather than only on ordinary reasoning. This type of rehearsed-in-advance consistency in the reasons presented goes beyond what can be expected of lay reasoners. Moreover, a lawyer is committed to defend her client come what may. On the other hand, it may be in the best interest of ordinary reasoners who, in spite of the myside bias, find or stumble upon counterarguments against their own views to take these counterarguments seriously and perhaps even to change their minds.

Most importantly, the lawyer analogy applies only to the *production* of arguments. During a court trial, each actor is ascribed a carefully defined role. To simplify things a bit, lawyers produce arguments, and judges and juries evaluate those arguments. To evaluate arguments, judges and juries also rely on reason, but they, unlike the lawyers, are not supposed to be biased. We will see in Chapter 12 how much the evaluation of arguments fits with the idealized picture of the disinterested judge.

In an adversarial trial, the two battling parties are locked in a zero-sum game: one side's win is the other side's loss. While this highlights the utility of the myside bias, it might also unnecessarily tie it to competitive contexts. In fact, even when people have a common stake in finding a good solution and are therefore engaged in a positive-sum game, having a myside bias may still be the best way to proceed.

Imagine two engineers who have to come up with the best design for a bridge. Whichever design is chosen, they will supervise the construction together—all they want is to build a good bridge. Ella favors a suspension bridge, Dick a cantilever bridge. One way to proceed would be for each of them to exhaustively look at the pros and cons of both options, weigh them,

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

and rate them. They would then just have to average their ratings—no discussion needed, but a lot of research.

Alternatively, they can each build a case for their favored option. Ella would look for the pros of the suspension bridge and the cons of the cantilever; Dick would do the opposite. They would then debate which option is best, listening to and evaluating each other's arguments. To the extent that it is easier to evaluate arguments presented to you than to find them yourself, this option means less work for the same result: Ella and Dick each have to find only half as many arguments to thoroughly review the pros and cons of each option.

The myside bias doesn't turn argumentation into a purely competitive endeavor. Argumentation is a form of communication and is typically pursued cooperatively. At its best, the myside bias becomes a way of dividing cognitive labor. In Chapter 12, we will see a similar dynamic at play in the way reason evaluates, rather than looks for, arguments.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

12

Quality Control: How We Evaluate Arguments

The Cicero we met in Chapter 11 recommending complete one-sidedness was a mere "boy," whose advice was "rough and incomplete." Many years later the Roman orator, rich of the "experience which [he] gained from so many and such important causes as [he has] pleaded,"[1] tells a different tale:

> If ever a person shall arise who shall have abilities to deliver opinions on both sides of a question on all subjects, after the manner of Aristotle, and, from a knowledge of the precepts of that philosopher, to deliver two contradictory orations on every conceivable topic, . . . and [who] shall unite with those powers rhetorical skill, and practice and exercise in speaking, he will be the true, the perfect, the only orator.[2]

Accumulating arguments for one side had become child's play for this veteran of many senate fights. To make a great case, an orator must combine rhetorical skills with the ability to anticipate and take into account potential counterarguments. While a speaker might still only present arguments for her side, her mind must be agile enough not to be blinded by the myside bias.

We claim that the myside bias makes sense when the goal is to justify one's ideas and actions or to convince others. But wouldn't justification or conviction be better served if reason allowed us also to find arguments against our side, even if only to refute them? A skilled orator spends time honing her argument—its content, its form, its delivery. From Cicero and Quintilian to contemporary speechwriters and spin doctors, massive efforts are expended

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

to devise convincing arguments. If reasoning, as we claim, evolved to change others' minds, shouldn't it look for such well-crafted arguments?

Here, then, is a potential problem for our interactionist theory: a substantial amount of evidence—and a passing familiarity with Internet comments—shows that people are far from being natural Ciceros. In Chapter 11 we encountered Deanna Kuhn's study of argumentation demonstrating the difficulty most people have in finding counterarguments. The same study showed that even in support of their own point of view, people often give rather weak arguments. Asked about her opinion on school failure, one of Kuhn's participants identified poor nutrition as the main culprit. Here's an excerpt from the interview:

Experimenter: If you were trying to convince someone else that your view is right, what evidence would you give to try to show this?
Participant: The points that they get in school. The grades that they get in school to show . . .
Experimenter: What would that show?
Participant: That they are lacking something in their body. That the kids who were failing lack something in their body.[3]

As Kuhn points out, the participant "makes it clear that the existence of the phenomenon itself is sufficient evidence that it is produced by the cause the [participant] invokes."[4] Other participants offered explanations that were close to a restatement of their theory. Most could not think of what would constitute evidence supporting their ideas. Kuhn's bleak assessment of untrained participants' argumentative skills is shared by other eminent psychologists, such as Richard Nisbett and Lee Ross, who suggest that people are generally content with the first reason they stumble upon,[5] or David Perkins, who asserts that many arguments make only "superficial sense."[6]

That reasoning shouldn't always be able to home in on excellent, knockdown arguments is obvious enough: looking for good reasons is a costly, time-consuming business. There has to be a trade-off between the quality of the reasons and the effort put in finding them. However, what these psychologists claim is not that we find very good, even if imperfect, reasons but that we find quite superficial, weak reasons.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

This is another problem for the intellectualist approach. If reason's function is to improve the lone reasoner's beliefs and decisions, it had better provide good reasons for them. In fact, however, people's criteria for their own reasons are pretty lax. So not only do people mostly find reasons that support their intuitions (the myside bias), they don't even make sure that the reasons are much good. It's hardly surprising that reason should by and large fail at correcting mistaken intuitions.

What about reason aimed at changing others' minds? Wouldn't people be more persuasive if they acted like lawyers, investing time and effort to anticipate potential counterarguments and find better arguments? Reason evolved to be used not in law courts but in informal contexts. Not only are the stakes of an everyday discussion much lower, but its course is also very different. A discussion is interactive: instead of two long, elaborate pleas, people exchange many short arguments. This back-and-forth makes it possible to reach good arguments without having to work so hard.

Interactive Argumentation Is Easier

Sociolinguists stress how much interlocutors help each other communicate effectively, for instance by providing constant feedback that they follow what the speaker is saying: all these "Hm hms," "Yeahs," nods, and so forth. Drawing on this tradition, the anthropologist and linguist Steven Levinson has argued that humans are endowed with an "interaction engine."[7] Our communicative abilities are tailored to the interactive context in which they naturally function. For instance, when we want to refer to someone, we often have a wide array of options: "Ms. Catherine Turk," "Ms. Turk," "Kate," "the head of the accounting department," and so forth. Which option is most appropriate in a given context depends, inter alia, on how well the interlocutor knows the person. Happily, if we use the wrong option, the miscommunication can easily be fixed:[8]

Michael: I had lunch with Kate.
Rob: Who?
Michael: Kate Turk.
Rob: I just arrived here. I don't think I know her.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Michael: Sorry. She's the head of the accounting department.
Rob: Ah, right.

Here Michael starts off with the most conventional way of referring to a familiar person in the United States: the first name. Rob's answers allow him to refine his description until he reaches the appropriate level.

Without using feedback from the interaction partners, deciding the most appropriate form to use would involve much reflection:

Michael, reflecting: Everybody here knows Kate. But Rob just arrived. However, Kate always makes an effort to introduce herself to everybody when they arrive. Rob got here two days ago, so he should know her. But I haven't seen Kate yesterday or the day before, so perhaps she was sick and didn't come. I'd better specify who she is just in case.

The solution people adopt—starting out with what seems to be the best option and, if necessary, refining it with the help of feedback—is the most economical. For communicators, being "lazy"—using the shortest form likely to be understood—is being smart.[9]

The feedback from the interlocutor is useful for at least two reasons. One reason is that it is the interlocutor's understanding that matters in the end, so it is easier to let her decide whether she understands or not. Another reason is that the interlocutor can do more than indicate she doesn't understand what the speaker means: she can actively guide the speaker's efforts, as Rob did when he told Michael why he doesn't know who Kate is.

Reasoners face a similar challenge: figuring out what is the best way to get their message across. They, too, could engage in elaborate guesswork to find the best argument in every situation. This strategy is both time and effort intensive, and far from being foolproof, as illustrated by the following exchange:

Sherlock Holmes meets his friend Watson at a coffee shop.

Holmes: My dear Watson, I didn't dare interrupt—you were in such charming company! I've been observing the two of you for a few minutes, and I must absolutely advise you to see this woman again!

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

> She's a perfect match for you. You may say that she's younger than you, but the age difference is not that large. I know you have a thing for brunettes, and she's blond, but that should easily be overlooked. I also observed that while you were talking of personal matters, there were no signs of physical intimacy, but I'm persuaded that would come quickly enough. Watson, I'm sure this woman is perfect for you!
>
> *Watson:* She's my sister.

Of course, Conan Doyle always rigged the situation so that his Holmes would not commit this kind of faux pas. Our Holmes, who is also fictional but a tad more realistic, was unable to guess that Watson's charming companion was his sister. Had he paused after exclaiming, "You were in such charming company!," Holmes would have been told that the lady was Watson's sister and would have avoided further embarrassment.

Feedback plays an important role even in simple forms of argumentation. Take this banal exchange:

> *Hélène:* We should go to Isami; it's a good restaurant.
>
> *Marjorie:* I don't know. I had Japanese last week already.
>
> *Hélène:* But this one is very original.

Isami might be a great place for many reasons—its originality, but also the prices, the freshness of the fish, the crowd—but Hélène doesn't list them all at the outset. Instead, she offers a generic summary assessment: "it's a good restaurant." This first argument doesn't sway Marjorie, but she doesn't simply say "no." Instead, she provides a reason for her dissent: "I had Japanese last week already." Thanks to Marjorie's feedback, Hélène can tailor her next argument, pointing out the quality of Isami that is most likely to change Marjorie's mind. In another context, the exchange might have gone as follows:

> *Hélène:* We should go to Isami; it's a good restaurant.
>
> *Marjorie:* I don't know. I don't have much money at the moment, and Japanese restaurants can be quite pricy.
>
> *Hélène:* But this one is quite cheap.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Among all the arguments Hélène could have put forward at the outset, some are much more likely to convince Marjorie than others. Hélène could have tried to anticipate which arguments would be most convincing, but that would have taken some effort—even in the unlikely event that she had access to all the pertinent information, from where Marjorie ate last week to the state of her bank account.

Admittedly, sometimes achieving prompt conviction is crucial. When Voltaire, the high priest of French Enlightenment, was about to be lynched by an English mob, he had to convince them quickly that he was a genuine anglophile (which he apparently did). Fortunately, very few discussions have such urgency. Most aim at deciding who should do the dishes, whether the Joneses should be invited for dinner, or what is the best movie in the theaters at the moment. Failing to immediately prevail in mundane discussions is nearly costless. Even when the stakes are higher—does the customer buy the car, does the policeman accept your account, should your family move to Singapore or stay in Hong Kong—failing with the first reason is rarely critical; more reasons can be tried out.

Providing a stream of poor reasons does carry a cost: making the speaker look daft. As a result, even casual arguers should exhibit a moderate degree of quality control regarding the reasons they provide. What is clear, however, is that our interactionist theory does not predict that humans should be born Ciceros, weaving complex arguments and spontaneously anticipating rebuttals. Reason should make the best of the interactive nature of dialogue, refining justifications and arguments with the help of the interlocutors' feedback.

Refining Reasons

The experiments presented earlier, which prompted psychologists to deplore the poor quality of the reasons put forward by participants, did not take place in a typical dialogic context. When a normal interlocutor is not swayed by a reason, she offers counterarguments, pushing the speaker to provide better reasons. An experimenter, by contrast, remains neutral. She may prompt the participant for more arguments, but she doesn't argue back. If reason evolved to function in an interactive back-and-forth, strong arguments should be expected only when they are called for by an equally strong pushback.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

To evaluate people's ability to construct arguments in a genuine argumentative setting, Lauren Resnick and her colleagues asked small groups of students to discuss a controversial topic. Participants were able to exchange several arguments every minute, bringing new ideas to the table and criticizing each other's suggestions and arguments. Thanks to this back-and-forth, the participants performed well and the researchers were "impressed by the coherence of the reasoning displayed. Participants . . . appear to build complex argument and attack structure. People appear to be capable of recognizing these structures and of effectively attacking their individual components as well as the argument as a whole."[10] Deanna Kuhn and her colleagues reached a similar conclusion in a more quantitative study:[11] they found that the students were able to produce much better arguments—about, say, the death penalty—after arguing about that topic with their peers. They clarified the links between premises and conclusions, added evidence to support their opinions, and relied on a wider range of argument types.

At the end of Chapter 11, we saw how the myside bias could turn into an efficient way of dividing cognitive labor, with each individual finding arguments for her side and evaluating arguments for the other side. The process we have just described is another elegant division of cognitive labor. Instead of engaging in a costly and potentially fruitless search for a knockdown argument, reasoners rely on the interlocutors' feedback, tailoring their arguments to the specific objections raised.

Fallacy? What Fallacy?

People exercise low quality control on their own reasons and are easily satisfied with relatively weak justifications and arguments. From the interactionist point of view, this is readily explained by the fact that reason evolved to work in an interactive setting. As predicted, people end up formulating better, more pointed arguments in the back-and-forth of a dialogue than when reasoning on their own.

What about quality control on other people's reasons, though? Here the interactionist approach makes very different predictions. If one can afford to be lax when evaluating one's own reasons, one ought to be more demanding when evaluating others' reasons. Otherwise we would accept the

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

silliest excuses as good justifications and the most blatant fallacies as good arguments. We would be all too easily manipulated. This prediction goes against a common idea that people are easily gulled by sophistry. This common idea, however, is wrong, in part because it relies on misguided criteria for evaluating arguments. When more sensible criteria are used, experiments reveal that people tend to accept reasons when they should, and only when they should.

A common criterion used to distinguish good from bad arguments is whether they can be categorized as an informal fallacy, from the *ad populum* to the *ad hominem.* Lists of such fallacies were already produced in classical antiquity and today can be easily found online. The issue is that for almost each and every type of fallacy in such lists, there are counterexamples in the form of arguments that meet the criteria to be declared fallacious but that in real life are quite acceptable or even good arguments, arguments that might convince a rational audience.[12]

Here is a *tu quoque* ("you too") fallacy:

Yoshi: You shouldn't drink since you'll be driving!
Makiko: Weren't you yourself caught driving under the influence a month ago?

Makiko's objection is supposed to be a fallacious argument against the advice given by Yoshi. After all, the fact that he does not follow his own advice does not make it wrong. On the other hand, if one suspects the speaker has a good reason not follow his own advice, then the *tu quoque* argument would be quite reasonable:

Yoshi: You shouldn't eat these chocolates that Aunt Hélène brought us; they are not very good!
Makiko: Didn't you almost finish the box?

In this case, Makiko's objection does cast a reasonable doubt on the reliability of Yoshi's advice.

Here is an *ad ignorantiam* fallacy (arguing that a claim is true because it is not known to be false):

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

The policeman: I am convinced that Ishii is a spy. I could find no evidence that he is not.

The policeman's argument is supposed to be fallacious because not knowing that a proposition is false is generally not a good argument that it is true. On the other hand, there are cases where such an argument is indeed quite good:

The policeman: I am convinced that Ishii is a law-abiding citizen. I could find no evidence that he is not.

We could multiply the examples but our point would each time be the same: *tu quoque* is fallacious except when it is not; *ad ignorantiam* is fallacious except when it is not; in fact, most if not all fallacies on the list are fallacious except when they are not. This is often implicitly acknowledged when the fallacies are given a more careful definition. The *tu quoque* fallacy is an *inappropriate* use of the fact that the interlocutor does not abide by her own judgment, the *ad ignorantiam* fallacy is an *inappropriate* use of the fact that the claim examined is not known to be false, and so on.

This way of defining fallacies gives us license to invent indefinitely many new types: the Mount Everest fallacy (inappropriate comparison to the Mount Everest in an argument), the chicken soup fallacy (inappropriate use of facts about chicken soup in an argument), and so on. After all, the political philosopher Leo Strauss invented the *reductio ad Hitlerum* fallacy, which consists in arguing against an opinion by inappropriately comparing it to a view that Adolf Hitler might have held. Fallacious references to Hitler are much more frequent than fallacious references to Mount Everest, so why not enjoy the fun of labeling such references with a special name?

The standard view of informal fallacies has been attacked by psychologists Ulrike Hahn, Mike Oaksford, and their colleagues.[13] They have detailed the variables that should make some arguments more convincing than others. For instance, in the appeal to ignorance, one of the main variables is: How likely are we to find positive evidence if we look for it? The spy argument is weak because it is difficult, even for a policeman, to find evidence that someone is *not* a spy—if spies were easy to recognize, they would not remain spies for long. By contrast, if a policeman looks for evidence that an indi-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

vidual is a delinquent when such is indeed the case, he is quite likely to find the evidence. If he does not, this is in fact evidence that the individual is not a delinquent. Moreover, spies are, fortunately, much rarer than good citizens, further undermining the spy argument.

Hahn and Oaksford's case is not only theoretical. They have tested their hypotheses by manipulating relevant variables and asking people to evaluate the resulting arguments. What they found, with a variety of examples, is that people rate arguments in a rational way and don't easily fall prey to genuinely fallacious arguments.

For instance, here's one of the arguments—an *ad ignorantiam*—that they asked participants to rate:

> This drug is likely to have no side effects because five meticulously controlled, large-scale clinical trials have failed to find any side effect.

When given this argument, people rate it as being quite strong. Participants also react appropriately when the argument changes. For instance, they are less convinced if only two trials failed to reveal any side effect. Similar results were obtained for different types of arguments. On the whole, the evidence shows that when people are presented with truly fallacious arguments, they are reasonably good at rejecting them.[14]

Evaluating One's Own Reasons as if They Were Someone Else's

The experiments we reviewed suggest that there is an asymmetry between how people produce reasons—they are relatively lax about quality control—and how they evaluate others' reasons—they are much more demanding. With Emmanuel Trouche, Petter Johansson, and Lars Hall, Hugo conducted a tricky experiment that aimed at making this asymmetry as plain as possible. It involved getting people to evaluate their own arguments as if they were someone else's.[15]

In the first phase of the experiment, participants tackled five simple reasoning problems regarding the products sold in a fruit and vegetable shop. For instance, they might be told that in the shop, "none of the apples are organic." From this, they had to draw a conclusion as quickly and intuitively

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

as possible, choosing among several options such as "Some fruits are not organic" and "We cannot tell anything for sure about whether fruits are organic in this shop." One participant whom we will call Rawan, for instance, selected as the correct conclusion, "Some fruits are not organic."

In the second phase, participants were asked to give reasons for their intuitive answers to each of the five problems they had answered in the first phase of the experiment. As they did so, they could, if they wanted, modify their initial answer. Given what we know about the production of reasons, we shouldn't expect much to happen at this juncture. Most participants should produce reasons that support their intuition without carefully checking how good their reasons are and without revising their initial selection. Indeed, only 14 percent of participants changed their minds, and the change was as likely to be for the better as for the worse. Rawan was among those who didn't change their minds. For her answer to the organic fruit problem, she offered the following justification: "Because none of the apples are organic, and an apple is one type of fruit, we can say that some of the fruits in the store are not organic."

In the third phase of the experiment, participants were given one by one the same five problems, and reminded of the answers they had given. They were then told about another participant who, on an earlier day, had answered differently, and they were given this participant's answer and argument. On the basis of this argument, they could change their mind and accept the other participant's answer or they could stick to their original answer.

With one of the five problems, we played a trick on the participants and told them that their answer had been different from what it had actually been. We told Rawan, for example, that she had answered, "We cannot tell anything for sure about whether fruits are organic in this shop." Moreover, we told her that someone else had selected the conclusion, "Some fruits are not organic" and given as justification "Because none of the apples are organic, and an apple is one type of fruit, we can say that some of the fruits in the store are not organic" (which had been Rawan's own selection and argument).

We hoped to have the perfect setup to test the asymmetry between the production and the evaluation of reasons: participants were led to evaluate an argument they had given a few minutes earlier as if it was someone else's.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

And it worked. About half of the participants, Rawan included, did not notice that they had been tricked into thinking their own reason was somebody else's.

Would participants we had successfully misled at least agree with the argument they had themselves produced a moment before? Well, no. Even though they had deemed the argument good enough to be produced, they became much more critical of it when they thought it was someone else's, and more than half of the time, they found it wanting. Reassuringly, there was a tendency for participants to be more likely to reject their own bad reasons than their own good reasons. Rawan, who had initially given a good reason, found the same reason convincing when she thought someone else had given it.

Accepting Good Arguments

So far, we have stressed one important feature of reason evaluation: it should be demanding enough to reject poor reasons. But it is just as important that it should accept good reasons. Reasoning, we argued in Chapter 10, serves a function both for communicators and for their audiences. In a situation where a communicator wants to make a claim that the audience is unlikely to accept just on her authority, reasoning generates arguments that the audience might evaluate and accept, and hence accept the claim. For the audience, reasoning is a tool of epistemic vigilance. It serves to evaluate arguments provided by a communicator so as to reject claims that are poorly supported and to accept claims that are well supported. Indeed, the whole point of epistemic vigilance is not just to reject dubious information but also to accept good information. For this we must be able to change our minds when presented with good enough reasons to do so.

To test this prediction, with Emmanuel Trouche and Jing Shao we conducted a series of experiments using the following problem, which we can call the Paul and Linda problem:[16]

> Paul is looking at Linda and Linda is looking at John. Paul is married but John is not. Is a person who is married looking at a person who is not married?

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

The three possible answers are "Yes," "No," and "Cannot be determined." Think about it for a minute—it's a fun one.

Most people answer "Cannot be determined," thinking that knowing whether Linda is married or not is necessary to answer the question for sure. But consider this argument:

> Linda is either married or not married. If she is married, then she is looking at John, who is not married, so the answer is "Yes." If she is not married, then Paul, who is married, is looking at her, so the answer is "Yes" again. So the answer is always "Yes."

If you are anything like our participants (Americans and Chinese recruited online), then you are likely to accept the argument. When we gave the participants the argument, more than half changed their minds immediately. By contrast, if you had figured out the problem on your own and we had told you, "The answer is 'Cannot be determined' because we don't know whether Linda is married or not," then you would never have changed your mind. The way people evaluate these arguments is remarkably robust.

When we gave our participants the argument for the correct answer, we didn't tell them it was *our* argument (as experimenters). We told them it had been given by another participant in a previous experiment. So they had no reason to trust her more than they trusted themselves. To make things worse, we told some participants that the individual who gave them the argument was really, really bad at this kind of task. We told others that the individual who gave the argument would make some money if the participants got the problem wrong. So they expected an argument from someone who they thought was either stupid or out to trick them. And yet when they saw the argument, most accepted it. Indeed, although the participants had said that they did not trust the individual giving them the argument one bit, this lack of trust barely affected their likelihood of accepting the argument.

We also asked some participants to think hard about the problem and to justify their answers. A few of them did get it right. But most got it wrong, and because they had thought hard about it, they were really sure that their wrong answer was right. Most of them said that they were "extremely confident" or even "as confident as the things I'm most confident about." But that

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

didn't make them less likely to change their mind when confronted with the argument above than participants who had had their doubts. Even though they could have sworn that the conclusion was wrong, when they read the argument, they were swayed all the same.

The Two Faces of Reason

In this chapter and in Chapter 11, we have looked at how reasoning produces arguments and at how it evaluates the arguments of other people. The results can be summarized in Table 2.

The "production of reasons" row is really bad for the intellectualist approach. When people reason on their own, they mostly produce reasons that support their decisions or their preconceived ideas, and they don't bother to make sure that the reasons are strong. As we'll see in Chapter 13, this is a recipe for disaster: not only is the solitary use of reason unlikely to correct mistaken intuitions, but it might even make things worse.

The fact that people are good at evaluating others' reasons is the nail in the coffin of the intellectualist approach. It means that people have the ability to reason objectively, rejecting weak arguments and accepting strong ones, but that they do not use these skills on the reasons they produce. The apparent weaknesses of reason production are not cognitive failures; they are cognitive features.

This picture of reason fits with the predictions of the interactionist approach. People produce reasons as predicted by the theory: they find reasons for their side—a good thing if their goal is to change others' minds—and

Table 2 The two faces of reason

	Bias	Quality control
Production of reasons	*Biased:* people mostly produce reasons for their side	*Lazy:* people are not very exigent toward their own reasons
Evaluation of others' reasons	*Unbiased:* people accept even challenging reasons, if those reasons are strong enough	*Demanding:* people are convinced only by good enough reasons

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

they start out not with the strongest reasons but with reasons that are easier to find, thus making the best of the feedback provided by dialogic settings. People also evaluate others' reasons as expected, rejecting weak reasons but accepting strong enough reasons, even if that means revising strong beliefs or paying attention to sources that don't inspire trust.

If we take an interactionist perspective, the traits of argument production typically seen as flaws become elegant ways to divide cognitive labor. The most difficult task, finding good reasons, is made easier by the myside bias and by sensible laziness. The myside bias makes reasoners focus on just one side of the issue rather than having to figure out on their own how to adopt everyone's perspective. Laziness lets reason stop looking for better reasons when it has found an acceptable one. The interlocutor, if not convinced, will look for a counterargument, helping the speaker produce more pointed reasons. By using bias and laziness to its advantage, the exchange of reasons offers an elegant, cost-effective way to solve a disagreement.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

13

The Dark Side of Reason

Arthur Conan Doyle's novel *The Hound of the Baskervilles* begins when a Dr. Mortimer tries to hire Sherlock Holmes's services:

> "I came to you, Mr. Holmes, because . . . I am suddenly confronted with a most serious and extraordinary problem. Recognizing, as I do, that you are the second highest expert in Europe ———"
>
> "Indeed, sir! May I inquire who has the honour to be the first?" asked Holmes with some asperity.
>
> "To the man of precisely scientific mind the work of Monsieur Bertillon must always appeal strongly."
>
> "Then had you not better consult him?"[1]

Who was this Monsieur Bertillon, superior even to Holmes as an expert detective? Alphonse Bertillon was, at the end of the nineteenth century, one of the most respected policemen and forensic scientists of the world. He had developed a scientific method for recording the identity of criminals based on anthropometry: a precise measurement of various traits, from the size of the left foot to the length of the right ear.[2] The most famous component of this system, still used today, is the mug shot, a standardized way of photographing people who have been arrested.

It is for his expertise in the new domain of photography that Bertillon is contacted by investigators from the French army in early October 1884. His task is to take a series of pictures of an extremely sensitive document, a letter known as the *bordereau*, written by a French officer spying for the Germans.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Bertillon obliges. A few years later, this tiny piece of paper will turn him into the most ridiculed man in France.

On October 13 Bertillon is asked to perform one more task, one that lies outside his area of expertise. In order to identify the *bordereau*'s writer, he must tell whether the writing on another letter matches that of the *bordereau.* After a few hours, Bertillon concludes that the same person very likely wrote both letters. That, however, is not conclusive enough for the lieutenant-colonels Henry and du Paty de Clam, who lead the inquest. They want a definitive indictment. And so, two days later, they visit Bertillon once again. Now they give him more information: they have evidence that Captain Albert Dreyfus is the spy who leaked French secrets to the Germans. All they need is a final proof: that the *bordereau* bears his handwriting. Actually Henry and du Paty de Clam have no other piece of evidence, but Bertillon does not know that, and with the extra confidence afforded by anti-Semitism, he happily starts working with the premises that the Jew Dreyfus is guilty.

Bertillon's mind works tirelessly with a single purpose: proving that Dreyfus wrote the *bordereau.* Here's what he has to work with: two letters—the *bordereau* and a sample of Dreyfus's writing—that have some similarities but also marked differences. These differences are sufficient for real experts to conclude that the two letters have not been written by the same person. But Bertillon is smarter than that. Only by imagining what clever deceptions Dreyfus has devised will this connoisseur of the criminal mind be able to prove the traitor's guilt.

Bertillon wonders: What kind of spy would write such a compromising message in his own hand? (The real spy, as it turns out, but no one knows this yet.) In Bertillon's mind Dreyfus, a spy, and a Jew to boot, is too shrewd to make such a glaring mistake. He must have disguised his hand. This explains the differences between Dreyfus's normal writing and the *bordereau.*

But now Bertillon has another problem: How to account for the similarities? Why hasn't that shrewd spy simply used a completely different writing? To answer this question Bertillon comes up with his chef-d'œuvre, the keystone of his system: the theory of the auto-forgery.

Imagining what a shrewd spy might do, Bertillon realizes that transforming one's writing would work only if the potentially incriminating document were found in a nonincriminating place. Then Dreyfus could use the disparities to

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

claim that he was not the author of the *bordereau.* However, if the letter were discovered on Dreyfus's person or in his office, he could not simply claim that it wasn't his. Instead, this master of deception would have to say that he was being framed, that someone had planted the *bordereau.* But if someone were to try to frame Dreyfus, surely they would be careful to reproduce his handwriting. And so *Dreyfus set out to imitate his own handwriting*—he engaged in auto-forgery. Frank Blair, a Chicago lawyer writing at the time of the affair, offers a sarcastic summary of Bertillon's reasoning:

> In short the differences between Dreyfus's natural hand and that of the bordereau, admitted by Bertillon, were, according to him, artfully put in by Capt. Dreyfus to throw off suspicion; while the absolute similarities were put in to enable him, in a proper case, to claim they were traced from his own handwriting, and therefore, done by someone else.[3]

If Dreyfus's handwriting looks like that of the *bordereau,* he is guilty. If it does not, he is guilty. The blatant flaw in the reasoning doesn't deter Bertillon, who states in conclusion: "I have arrived at a set of observations and comments that embrace all the facts with a comprehensiveness so perfect that the conclusions impose themselves indisputably."[4] Partly on the strength of Bertillon's "evidence," Dreyfus is arrested, tried, and sentenced to cashiering and life deportation—in a minuscule cell, on a minuscule island, far away from France.

A year later, Lieutenant-Colonel Georges Picquart takes over as head of military intelligence. Reviewing the files associated with the Dreyfus case, Picquart discovers conclusive evidence that the real culprit is the officer Ferdinand Walsin Esterházy. His handwriting matches that of the *bordereau* perfectly. When Bertillon is presented with this evidence, even he has to admit that Esterházy's writing is strikingly similar to the writing on the *bordereau.* But Bertillon's mind never falls short of reasons to sustain his views; Bertillon claims that "the Jews have been training someone for a year to imitate the writing"—to become a scapegoat, presumably.[5] The higher echelons of the military are similarly unmoved. They already have their culprit; another one would just mess up their story. Thanks to Picquart, Esterházy is tried. Thanks to the generals' influence, the guilty man is proven innocent, the

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

innocent man proven guilty once more, and the whistle-blower disgraced. Dreyfus stays on his island.

It will take a major social and political upheaval, brought about by Jean Jaurès, Leon Blum, Georges Clemenceau, and, most famously, the novelist Emile Zola and his *J'accuse,* for the army to reopen the case. In 1899, the original decision is quashed and a new trial is held in Rennes. Bertillon, again, is one of the expert witnesses. His task has become more difficult, as he must now prove that the *bordereau* is not an undisguised note left by the sloppy spy Esterházy but the carefully designed product of Dreyfus's devious mind. Bertillon is up to the task. Apotheosis of Bertillon's system, the deposition runs for more than fifty pages of dense text, plus pictures and exhibits.

Of that *bordereau,* Bertillon peruses every word, measures every letter, photographs every wrinkle. He sees patterns everywhere. When the thirteenth line of the *bordereau* is superimposed on the thirtieth, three letters are aligned.[6] When the word *intérêt* is taken out and repeated and the two copies put end to end, they measure 12.5 millimeters, a unit size on military maps.[7] Even more damning, a standard subdivision of this unit, 1.25 millimeters, is found everywhere in the word: "length of the *t*'s cross: 3 [units of 1.25 millimeters]; length of the acute accent: 1; width of the circumflex, one and a half, and height of the final *t:* 4, etc."[8] Coincidences? Impossible. The *bordereau* must be the work of a master craftsman who used several templates and a military-issue ruler to create one of the most complex forgeries of modern times.

Such considerations can leave no place for doubt and so, after ten hours of deposition(!), Bertillon gives a forceful conclusion: "In the set of observations and concordances that form my demonstration there is no place for doubt, and it is made strong by a certainty both theoretical and material that, with the feeling of responsibility born of such an absolute conviction, I affirm with all my soul, today as in 1894, under oath, that the *bordereau* is the work of the accused. I am done."[9]

It is hard to tell how impressed the court is with Bertillon's arguments. In any case, the judges find Dreyfus guilty of treason once again, although with "mitigating circumstances." This nonsensical verdict reflects more the need to uphold the status quo than the merits (or lack thereof) of the case. Dreyfus's innocence is plain for anyone to see. Refusing to wait for yet a new trial that may never happen, Dreyfus consents to be pardoned by President Loubet

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

on September 19, 1899, at the cost of accepting the guilty verdict. He will have to wait seven more years for his final rehabilitation: being reinstated to his former rank in the army.

The Bertillon in All of Us

Bertillon offers a fascinating study in the use of reasoning to defend preexisting beliefs. He seems to have been truly convinced by his own arguments. When three experts were tasked with evaluating his system, they found an "incomprehensible jumble" "totally devoid of scientific value" whose absurdity was "so blatant that one is hard put to explain the length of its exposition."[10] Yet they also concluded that the very "naiveté with which [Bertillon] unveiled the secrets [of his system] would lead one to believe in his good faith."[11]

It would be easy to regard Bertillon as a madman—many of his contemporaries did. But that would overlook his otherwise successful professional career, how he rose through the ranks and devised new ways to catch criminals—hardly what you would expect of a lunatic. And lest we feel too smug, every aspect of Bertillon's thinking has been reproduced in the laboratory, showing how reasoning can lead everyone on the wrong track. These experiments have replicated—on a smaller scale, fortunately—the mental processes occurring in Bertillon's mind. Unambiguously, they point to reasoning as the culprit.

When Bertillon mentions the perfect comprehensiveness with which he embraces all the facts, he exhibits clear symptoms of overconfidence. According to the intellectualist approach, reasoning is supposed to make us doubt our own beliefs, especially when they rest on foundations as shaky as Bertillon's. How can reasoning lead to overconfidence instead?

Asher Koriat suggested an answer more than thirty years ago.[12] In one of his experiments, participants had to answer general knowledge questions, such as, "Does Corsica belong to France or Italy?" and to specify how confident they were. Participants were overconfident. If they thought, say, that they would be right in 80 percent of the cases on average, they might have the correct answer only six times out of ten.

While some participants were not given any special instructions, others were told to give reasons supporting their answer. This had no effect on

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

confidence because participants had been doing that all along: piling up reasons supporting their initial hunch. This is why they were overconfident. And it's not as if they were completely unable to think of reasons why they might be wrong—when asked to provide such reasons, they obliged, and became less overconfident. But the myside bias stopped them from spontaneously engaging in this more objective reasoning.

Between his first, not fully conclusive report and the deposition in Rennes fifteen years later, Bertillon has not only become even more confident, his beliefs have also undergone severe polarization. Dreyfus comes out more cunning, the conspiracy wider, Bertillon's testimony more critical than ever. Reason had had years to push Bertillon toward such extremes, but the roots of polarization can be observed in a much briefer time. When participants were made to think about someone for a few minutes, they ended up liking him more if they had liked him at first, and liking him less if they hadn't liked him at first.[13] In this brief interval of time, they had piled up reasons supporting and reinforcing their initial impression.

Initially, the only evidence Bertillon had of Dreyfus's guilt was the alleged resemblance between his handwriting and that of the *bordereau.* When Picquart showed Bertillon Esterházy's handwriting, which exactly matched the *bordereau,* Bertillon should have immediately changed his mind. But between his first encounter with the case and Picquart's intervention, Bertillon had built an unyielding scaffold of reasons. This scaffold upheld his original contention in spite of overwhelming evidence against it, a process known as belief perseverance.

The earliest experimental demonstration of belief perseverance was done by a team led by Lee Ross in 1975.[14] Participants were asked to distinguish between real and fake suicide notes, and were told how well they'd done. They were then left to think about their performance for a little while. During this time, they thought of many reasons why the feedback made sense: they had always been very sensitive, they knew someone who was depressed, and so on. Then the participants were told that in fact the feedback had been completely bogus, bearing no relationship whatsoever with their actual performance. But it was too late. Participants had found many reasons to buttress their beliefs, so that even when the initial support was removed, the beliefs stood on their own. Participants who had received a positive feedback thought

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

they had done better than those who had received negative feedback, even after they'd been told the feedback was bogus.

Reasoning Poorly Together

Bertillon offers a sad example of how "starting with a mistake, a remorseless logician can end up in bedlam."[15] He kept accumulating reasons for his initial belief in Dreyfus's culpability, with a blatant lack of self-criticism, until he reached grandiose and absurd conclusions. Bertillon's mind seems boundlessly fertile, but it is only one mind. Imagine what people can do when they gang up to find reasons supporting their beliefs.

Conspiracy theories often start small, questioning accepted facts: Why does the American flag allegedly planted on the moon appear on the photos to be waving in spite of the lack of wind? Could the World Trade Center really have crumbled on 9/11 the way people say it did? Some of the people who raise these questions go online, discover pamphlets, find kindred spirits. Soon enough the doubt escalates, alternative answers are found, and pointed questions turn into full-blown paranoia. Officials must be lying. The government has to be in on it—NASA, the CIA, the FBI, the NSA. The conspiracy has to be global, pushed by the United Nations, the Bilderberg group, and more often than not, some form of "international Jewry," heirs to the anti-Semitic conspiracy theories of Bertillon's days.[16]

While the development of vast conspiracy theories involves hundreds of people, smaller groups can also be led astray by reason. In the 1960s, the Yale psychologist Irving Janis started investigating when and why small groups make poor choices. He examined in detail the process that led to disastrous decisions such as the failed attack on Cuba launched by the American government in 1961—the infamous Bay of Pigs invasion. For Janis, the culprit was groupthink, the failure of group members to criticize each other's suggestions and to consider alternatives.[17]

The problems caused by a lack of dissent have also been captured in the laboratory, where psychologists have accumulated evidence of group polarization. Put a bunch of people together and ask them to talk about something they agree on, and some will come out with stronger beliefs. Racists become more racist, egalitarians more egalitarian.[18] Hawks increase

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

their support for the military; Doves decrease it.[19] When you agree with someone, you don't scrutinize her arguments very carefully—after all, you already accept her conclusion, so why bother? When like-minded people argue, all they do is provide each other with new reasons supporting already held beliefs. Just like solitary reasoners, groups of like-minded people can be victims of belief polarization, overconfidence, and belief perseverance.[20]

Evolution Doesn't Care How Good We Feel

Reasoning can lead to outlandish territories—Bertillon developing his system, a conspiracy theorist thinking that lizard men control the earth. It is only natural to dismiss people who adopt such beliefs as crazy or stupid—or, more politely, to suggest that they suffer from cognitive limitations.

The facts do not support this interpretation. Bertillon was neither a madman nor a dunce. Few conspiracy theorists suffer from psychosis or cognitive impairment.[21] Moreover, people who cannot be suspected of any mental deficiency share the same plight. Linus Pauling went from seeing vitamin C as a remedy for the common cold to hailing it as a universal cure. The escalation of the Vietnam War was decided by "extraordinarily intelligent, well-educated, informed, experienced, patriotic, and capable leaders."[22]

This spells trouble for the intellectualist approach. Not only does reasoning fail to fix mistaken intuitions, as this approach claims it should, but it makes people sure that they are right, whether they are right or wrong, and stick to their beliefs for no good reason. Historical examples attest that these are not minor quirks magnified by clever experiments, but real phenomena with tragic consequences.

Psychologists sometimes use a distinction between cognitive and motivational explanations. People who do or believe something wrong either must be the victims of a cognitive failure or must have been motivated to go astray. Since cognitive failures cannot explain the surprising outcomes of reasoning, a sensible move is to offer a motivational account, as Ziva Kunda did when she defended the prevalence of motivated reasoning. According to her, when reason leads people astray, it is because of a "wish, desire, or preference" to reach a preordained belief, whether it is accurate or not.[23] People may want

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

to believe something for many reasons, but the most common is hedonic: because it makes them feel good. There would be a desire to believe that leads people to "give preferential treatment to pleasant thoughts and memories over unpleasant ones."[24]

At its most extreme, a feel-good account could simply claim that people adopt whatever belief is pleasurable—an extreme form of wishful thinking. "Wishful thinking" is a phrase commonly used to describe someone's belief when it seems to be grounded not on evidence that the belief is true but on the desire that it were. Moreover, in ordinary conversation, "It's wishful thinking" is often given not just as description but as explanation: "Why does John believe he is popular? It's just a case of wishful thinking—he wishes he were!"

From a psychological point of view, and even more from an evolutionary one, wishful thinking is something to be explained rather than an explanation. Our beliefs are supposed to inform us about the world in order to guide our actions. When the world is not how we would want it to be, we had better be aware of the discrepancy so as to be able to do something about it. Thinking that things are the way one wishes they were just because one so wishes goes against the main function of belief.

It should therefore be no surprise that Kunda and other specialists of motivated reasoning have shown that people do not simply adopt beliefs as they see fit. They look for reasons, making sure that they can provide some justification for their opinions—and they drop even cherished beliefs when they fail to find justifications for them. For instance, people have an overall preference for believing they are better than average—smarter, better at socializing, more sensible, and so forth. However, they do not simply believe what would make them happiest: that they are the best at everything. Instead, they selectively self-enhance, only providing inflated assessments when they are somehow defensible.

For instance, people tend to think they are more intelligent than the average—that's an easy enough belief to defend: they can be good at math, or streetwise, or cultured, or socially skilled, and so on. By contrast, there aren't two ways of being, say, punctual. Since people can't think of ways to believe they are more punctual than the average, they just give up on this belief, or on other beliefs similarly hard to justify.[25]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Even if reasoning is not wishful thinking, its function could still be to make us feel good. A first problem with this hypothesis is that reasoning often has the opposite effect. Bertillon might have been engulfed by the pleasure of complete self-confidence, but he may also have been distressed by every new proof of Dreyfus's devilish ingenuity. The journalist Jonathan Kay interviewed many conspiracy theorists in the writing of *Among the Truthers* (people who do not accept the standard account of 9/11). He found people suffering from "debilitating emotional agony" caused by "sudden exposure to the magnitude of evil threatening the world."[26] If reasoning is supposed to make people feel good, it fails abysmally. Other cases are easy to conjure—the jealous husband who persuades himself his wife is cheating on him, the pessimist who keeps finding reasons why humanity is bound to self-destroy, the hypochondriac who looks for the symptoms of yet another disease.[27]

Not only does a feel-good explanation fail to fit the facts, it doesn't make evolutionary sense, either. It confuses the *proximal* and the *ultimate* levels of explanation.[28] A proximal explanation aims at pinpointing the psychological or neurological causes of a behavior. For instance, if Michael gets thirsty and drinks some water, the pleasure he derives from drinking could be a proximal explanation for his behavior: he drank the water because he anticipated that it would make him feel good.

Ultimate explanations, by contrast, answer questions at the evolutionary level. At the ultimate level, feeling good is no more than a means to an end. For evolution, hedonic states—pleasure, pain, happiness, despair—serve the purpose of motivating animals to perform certain actions critical to their survival and reproduction. We feel pleasure while quenching our thirst because drinking is necessary for survival. We experience pain when touching a burning log so that we withdraw our hand and avoid long-term damage. We like spending time with friends because having partners and allies has been crucial to reproductive success in human evolution. For the same reason, we despair if our friends abandon us. An individual who would find drinking painful but would enjoy feeling his hand roast or who would resent the affection of friends and revel in their loathing would not be well equipped to survive and reproduce. So whether or not reasoning helps people feel good, it cannot have evolved to this end.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Adaptive Lags in Reasoning

The interactionist approach can account for the various epistemic distortions introduced by reason—overconfidence, polarization, belief perseverance. Chapters 11 and 12 pointed out two major features of the production of reasons: it is biased—people overwhelmingly find reasons that support their previous beliefs—and it is lazy—people do not carefully scrutinize their own reasons. Combined, these two traits spell disaster for the lone reasoner. As she reasons, she finds more and more arguments for her views, most of them judged to be good enough. These reasons increase her confidence and lead her to extreme positions.

Many psychologists might agree with this diagnostic. However, such an explanation should only be a first step, soon followed by: Why on earth would reasoning behave that way? When an artifact fails to produce the desired results, this might be because it is broken, but it might also be because it is operating in abnormal conditions. If your pen doesn't work upside down, if your car doesn't start with an empty tank, it is not because they are out of order but because they are not designed to function in such conditions. Biological devices also have normal conditions: the conditions to which they are adapted.[29] The normal conditions for human lungs are formed by the earth's atmosphere around ground level. Our lungs work splendidly in these conditions, but less well or not at all in abnormal conditions—high altitudes, under water, in a tank full of helium, and so forth.

In our interactionist approach, the normal conditions for the use of reasoning are social, and more specifically dialogic. Outside of this environment, there is no guarantee that reasoning acts for the benefits of the reasoner. It might lead to epistemic distortions and poor decisions. This does not mean reasoning is broken, simply that it has been taken out of its normal conditions. In the same way, when objects take on new colors under the sodium lighting of an underground parking lot, our color perception is not broken; it is simply working in an abnormal environment. The artificial lights that have replaced the lighting we encountered during our evolution—chiefly, the sun—mislead our color perception.

This explanation—that reasoning now often works in an abnormal environment—is incomplete. If a bomb explodes inside the bomber plane rather

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

than when it hits the intended target, the engineer in charge does not get kudos by pointing out that the explosion was exactly of the predicted force. When the bomb explodes is at least as important as how it explodes. Similarly, when a cognitive mechanism is triggered is at least as important as how it works once triggered.

The basic trigger of reasoning is a clash of ideas with an interlocutor. This clash prompts us to try to construct arguments to convince the other or at least to defend one's own position. This trigger works also in the absence of an actual interlocutor, in anticipation of a possible disagreement. Sometimes this anticipation may be quite concrete: a meeting is already scheduled to try to resolve a disagreement or to debate opposing ideas. At other times, we might just anticipate a chance encounter with, say, a political opponent, and mentally prepare and rehearse arguments we would then be eager to use. There are even times when we replay debates that have already taken place and think, alas, too late of arguments that we should have used.

Sasha, for instance, is about to ask his mother to let him go to the all-night party at Vanessa's. He has been rehearsing his arguments: he has been working very well at school; his homework for the next week is done; the party will be a small affair, nothing wild, nothing his mother should worry about. The more he thinks about it, the more Sasha becomes convinced that his request is perfectly reasonable and that his mother should, of course, say yes.

Several things can happen then. Sasha might convince his mother that there are no serious objections to his going to the party. Or his mother might convince him that it is not such a good idea after all—she has heard from other parents that the party might be crashed by older kids who would bring beer and perhaps even drugs. Also, he seems to be forgetting that there is an exam next week for which he has not yet prepared. Listening to his mother's arguments, Sasha might want to argue back. At the end of the back-and-forth, either one will have convinced the other, or at least both will have given reasons to justify their points of view.

By contrast, if his mother just said no without paying attention to his argument, or if he never mustered the courage to ask, Sasha would probably see the arguments he never gave as compelling; he would see himself as a victim of parental injustice and incomprehension. Reasoning in anticipation of a discussion is fine—as long as the discussion actually takes place.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

What is problematic isn't solitary reasoning per se, but solitary reasoning that remains solitary. Reasoning, however, is bound to sometimes remain in one's head, as people cannot fully anticipate when they will be called to defend their opinions. Just as one can be taken aback by an unanticipated quest for justification, one can prepare for a confrontation of points of view that never materializes. The latter case may well be more common because of the difference in costs between the two types of failures. Being caught unprepared to defend an opinion or an action that others might object to is likely to be worse, and hence less common, than rehearsing a defense that in the end will not serve.

Modern environments distort our ability to anticipate disagreements. This is one of many cases in which the environment changed too quickly for natural selection to catch up. For example, our modern environments make some psychoactive substances, from coffee to cigarettes to alcohol, widely available. Some of these substances, such as cigarettes, are clearly bad for their users' fitness (in both meanings of the word). Yet we haven't evolved an innate disposition to avoid these substances in the same way that we have innate dispositions to avoid poisonous foods. Arguably, the explanation is that these substances would have been much rarer during our evolution and that they became common enough too recently for our brains to adapt to the change.

Have environmental changes thrown off-balance our ability to anticipate disagreements in the same way they made our reactions to psychoactive substances dangerous? Life in a modern, affluent society is different in myriad ways from life in the ancestral environment, and some of these differences are bound to affect the way we reason. For instance, before the invention of the printing press and the advent of modern media, people were typically made aware that somebody in their own group had opinions different from theirs thanks to interaction with that person. Finding out about difference of opinion and trying to resolve them commonly occurred through repeated exchanges of arguments that could be anticipated and mentally rehearsed. Nowadays we are inundated with the opinions of people we will never meet: editorialists, anchormen, bloggers. We are also expected to have an opinion on many different topics—from politics to music to food—and to be able to defend this opinion when challenged, giving us reasons to prepare for a variety of debates that might never occur.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

And this only scratches the surface of the problem. More dramatic changes affect the workings of reason. Big-city dwellers meet more strangers in any single day than their ancestors did in their lifetime. Many of these strangers have different cultural backgrounds. It is easy to see how this novel mix generates possibilities for disagreement, making it considerably more complex to properly anticipate the need for justifications.

Some cognitive mechanisms have been so fully repurposed by the modern world that they bear only a small resemblance to their ancestral form—witness the transformations brought by literacy to our ability to recognize simple arbitrary shapes.[30] While we do not believe that reason has undergone such dramatic transformations, environmental changes have certainly had an effect on when reason is triggered, on how it functions, and even on what goals it achieves. Reason is used now in a variety of ways that differ from its evolved function—from displaying one's smarts in a formal debate to uncovering the laws of physics. Unfortunately, some of those new uses of reason, such as preparing for debates that never come, turn out to be potentially harmful to the reasoner. As Keynes put it, "It is astonishing what foolish things one can temporarily believe if one thinks too long alone."[31]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

14

A Reason for Everything

In Chapter 13, when solitary uses of reason led people astray, it was because they started out with a strong intuition—that Dreyfus was guilty, that this was the right answer to the problem, and so on. The myside bias, coupled with lax evaluation criteria, make us pile up superficial reasons for our initial intuition, whether it is right or wrong. Often enough, however, we don't start with a strong intuition. On some topics, we have only weak intuitions or no intuitions at all—a common feeling at the supermarket, when faced with an aisle full of detergents or toilet papers. Or sometimes we have strong but conflicting intuitions—economics or biology? Allan or Peter? Staying at home with the kids or going back to work?

These should be good conditions for the individualist theory to shine. Reason has a perfect opportunity to act as an impartial arbiter. When the reasoner has no clear preconception, the myside bias is held at bay and reason can then guide the reasoner's choice, presumably for the better. Perhaps it is from such cases that beliefs about the efficiency of reason are born in the mind of philosophers. They make it a duty to examine precisely these cases in which intuitions are weak or conflicting. And if there is not enough conflict, philosophers excel at stirring it up: Are you sure other people exist? Think again! There is a philosophical theory, solipsism, that says other people don't exist or, at least, that you can never be sure that they do. Situations where intuitions are absent, weak, or conflicting might provide perfect examples of reason working in line with the expectations of the classical theory: reach a status quo between different intuitions, and only then let reason do its job. Let's look at

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

what reason does in such cases, starting with a clever experiment conducted by Lars Hall, Petter Johansson, and Thomas Strandberg.

As you walk along the street, a young man approaches you with a clipboard and asks whether you would be willing to take part in a short survey. For once, you accept. He hands you the clipboard with two pages of statements on political, moral, and social issues such as "If an action might harm the innocent, it is morally reprehensible to perform it." You must indicate how you feel about each statement on a scale going from "Completely disagree" to "Completely agree." You fill in the survey and hand the clipboard back. You're not quite done, though: the young man passes the clipboard back and asks you to explain some of your ratings. You happily do so—after all, you take pride in being an informed, thoughtful citizen with sound opinions.

What you haven't realized is that in the few seconds during which he held the clipboard, the young man—who is in fact an experimenter—has, by means of a simple trick, replaced some of the statements on the page with statements having the exactly opposite meaning. For instance, the statement about harming the innocent would now read, "If an action might harm the innocent, it is morally permissible to perform it" (with "permissible" having replaced "reprehensible") If some statements have been flipped, the answers haven't, so that for these statements, the sheet now indicates that you hold the exact opposite of the opinion you asserted one minute earlier. If you had indicated that you strongly agreed with the first statement, the sheet now says that you strongly agree with the second statement, which means the opposite.

Fewer than half of the participants noticed that something was wrong with the new answers. The majority went on justifying positions contrary to those they had professed a few minutes earlier, especially if their opinions weren't too strong to start with.[1]

Our boundless ability to produce reasons for just about anything we believe (or we think we believe, as shown by Hall and colleagues) has become a staple of social psychology since the pioneering experiments of Richard Nisbett and Tim Wilson in the 1970s that we evoked in Chapter 7. In one experiment, Nisbett and Wilson were standing outside malls pretending to sell stockings.[2] Some passersby stopped at their stall, made a choice, and, when asked, happily justified their decision: "This one looks more resistant"; "I prefer the color of that one." But the psychologists knew all these explana-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

tions to be bogus: they had mischievously displayed strictly identical pairs of stockings. That all the stockings were the same did not stop people from expressing preferences, which must have been based, then, on the position of the pairs in the display (many participants showed a right-most bias, for instance).

In such experiments participants start out with weak intuitions. In the substitution of statements experiment, it is mostly those participants who had expressed only mild agreement or disagreement and who therefore didn't have strong intuition on the issue who failed to detect the manipulation. In the second experiment, the stockings were all the same, so whatever preference was created by their position would have been very weak. Still, reason doesn't do the job classically assigned to it. It does not objectively assess the situation in order to guide the reasoner toward sounder decisions. Instead, it just finds reasons for whatever intuition happens to be a little bit stronger than the others. Humans are rationalization machines. As Benjamin Franklin put it, "So convenient it is to be a reasonable creature, since it enables one to find or make a reason for everything one has a mind to do."[3]

There are cases, however, where reason has a demonstrable impact on people's decision—but not one that fits with the intellectualist approach.

When Reasoning Makes a Difference

We find Tim Wilson again, except that this time he is dealing in posters, not stockings, and there is no trick: all the posters are different. The experiment is straightforward. Some participants are asked to rate five posters, period. Others have to rate the same five posters, but also to explain their ratings.[4] Being asked to reason affected participants' choices, such as by making them give higher ratings to humorous posters.

Many other studies have demonstrated that reason can make a difference. Some experiments require people to justify their decisions;[5] others give participants some extra time to reflect on their choices;[6] still others pit decisions based on feelings against decisions based on reasoning.[7] Each time, people who reason more act differently from those who reason less or not at all. A mere rationalization machine is not supposed to influence decisions. What is happening? Is reason helping people make better decisions?

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Not quite. In every one of these experiments, more reasoning led to worse decisions. For instance, in Tim Wilson's experiment, the participants were given the poster they had ranked higher to take home. Asked a few weeks later about their appreciation of the poster, those who had had to explain their preferences were less satisfied than those who had relied on their unfiltered intuitions.

To understand why reason can mess up people's decisions even when the myside bias is not the main culprit, we must look more precisely at how reason affects decisions.

Itamar Simonson performed an early experiment on this topic.[8] He started by designing two products—for example, beers—that would be equally preferred by most people. Let's call the first brand Beer Deluxe. It's a fancy product, with a quality rating of 75 out of 100, worth $18 for a six-pack. The second is Beeros, a less sophisticated—rating at 65—but cheaper—$12—alternative. When people had to choose between these two brands, they were indifferent, picking either beer about as often. Then the experimenter introduced a third brand, Premium Beer. At $18, Premium Beer is as expensive as Beer Deluxe, but it is also less well rated—a 70 out of 100. Given that Premium Beer is simply inferior to Beer Deluxe, it should not make a difference in people's choices. In fact, it did: once Premium Beer was introduced, people were more likely to pick Beer Deluxe.

Christopher Hsee conducted one of the most original experiments in the area. He asked participants which of two treats they would prefer to receive as a gift for having completed a task. Both gifts were chocolates, but one was a small (0.5 ounce), cheap ($0.50), heart-shaped chocolate while the other was a big (2 ounces), expensive ($2), roach-shaped chocolate. When participants relied more on their feeling, they were about split between the two choices. But when they reasoned to make a decision, most picked the big roach-shaped chocolate.[9]

Debora Thompson explored the phenomenon of feature creep: the multiplication of useless features that burdens so many gadgets and, in the end, reduces their usability. With her colleague Michael Norton, they showed that when people feel they must provide reasons for their decisions, they are more likely to pick a feature-rich item—such as a digital video player with dozens of functions—even though they realize it would be less convenient to use.[10]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Here's the common thread in all these results: in each case, reason drives participants toward the decision *that is easier to justify*. "Beer Deluxe is better but not more expensive than Premium Beer, so I'll pick it." "Given it's a gift, it would be irrational not to pick the bigger and more expensive chocolate just because of its shape—it's not as if it was a real roach anyway." "Why buy a digital video player that does fewer things?"

This common phenomenon is known as reason-based choice: when people have weak or conflicting intuitions, reason drives them toward the decision for which it is easiest to find reasons—the decisions that they can best justify.

Paying for Reasons

While these results are difficult to reconcile with the intellectualist theory—reason should lead people to better decisions, not to worse decisions—they are what the interactionist approach predicts. Reason doesn't stop being a social device in the absence of a point of view to uphold. Instead, it samples potential reasons for the different options available and drives the reasoner toward the decision that is the easiest to justify—whether or not it is otherwise a good decision.

In many cases, it looks as if reasoning is driving people toward worse, less rational decisions. The introduction of an obviously inferior option—Premium Beer—should not influence the decision between two superior options. Psychologists studying disgust can tell you that people will not enjoy eating that roach shaped chocolate, however big.[11] A gadget bloated with useless features will become a source of anxiety, not enjoyment.[12] Refusing to buy a jam simply because there are more jams to pick from doesn't make much sense.

Even more strikingly, people are willing to pay simply to have a reason for their decision. Amos Tversky and Eldar Shafir, who were among the first, with Itamar Simonson, to explore reason-based choice, asked a first group of participants to imagine the following scenario:

> You have just taken a tough qualifying examination. It is the end of the fall quarter, you feel tired and run-down, and you are not sure that you passed the exam. In case you failed, you have to take the exam again in a couple of months—after the Christmas holidays. You now have an

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

opportunity to buy a very attractive 5-day Christmas vacation package in Hawaii at an exceptionally low price. The special offer expires tomorrow, while the exam grade will not be available until the following day.

Would you

x. Buy the vacation package.
y. Not buy the vacation package.
z. Pay a $5 nonrefundable fee in order to retain the rights to buy the vacation package at the same exceptional price the day after tomorrow—after you find out whether or not you passed the exam.[13]

A second group of participants was asked to imagine that they had passed the exam and a third group that they had failed.

Most of the participants told that they had passed the exam decided to buy the vacation package—they reasoned that it was a well-deserved reward for their success. Most of the participants told that they had failed the exam also decided to buy the vacation package—they reasoned that they direly needed a break to recover from this failure.

Combined, these two results imply that it would be rational for most participants in the first group, who didn't yet know whether they had passed or failed, to buy the package and not to waste five dollars to postpone their decision. Whether they passed or failed, they would buy it. However, participants in this group chose to pay the fee and to wait a couple of days in order to know the exam results. Their problem: the reasons for buying the package were incompatible, one being "I deserve a reward for success" and the other "I need a break after failure." And so they paid to wait, effectively buying a reason to make a decision that they would make either way.

Social Rationality

When assessing decisions, it might seem clear that we should focus on the fit between the content of our decisions and our practical goals. We should buy posters we will enjoy more. When buying electronic devices, we should pick a model that best meets our needs. However, if being rational is striving for

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

the best decision, all things considered (and not just our practical goals), then making a good decision gets more complex.

Humans constantly evaluate one another. Are the people we interact with competent and reliable? Is their judgment sound? As we argued in Chapter 7, much of this evaluation is done in terms of reasons: we understand others' ideas and actions by attributing to them reasons, we evaluate the goodness of these reasons, and we evaluate people's reliability on the basis of their reasons. The way we rely on reasons may distort and exaggerate their role in thought and action, but it is not as if a better, more objective understanding were readily available. After all, psychologists themselves are still striving to develop such an understanding, and they disagree as to what it would look like. Reason-based understanding, for all its shortcomings, has the advantage of addressing two main concerns: providing a basis for an evaluation of people's performance, and providing an idiom to express, share, and discuss these evaluations.

Just as we evaluate others, they evaluate us. It is important to us that they should form a good opinion of us: this will make them more willing to cooperate and less inclined to act against us. Given this, it is desirable to act efficiently not only in order to attain our goals but also in order to secure a good reputation. Our reasons for acting the way we do shouldn't just be good reasons; they should be reasons that are easily recognized as good.

In some situations, our best personal reasons might be too complicated, or they might go against common wisdom, and hence be detrimental to our reputation. In such a case, it may be more advantageous to make a less-than-optimal choice that is easier to justify than an optimal choice that will be seen as incompetent. We might lose in terms of the practical payoff but score social points yielding a higher overall payoff.

Reason influences our decisions in the direction of reputational gains. For instance, those participants who picked Beer Deluxe because it was the easiest decision to justify may not have maximized their product satisfaction, but they scored social points: their decision was the least likely to be criticized by others.[14] Customers who ended up with a device burdened with useless features are (ironically) regarded as technologically savvy.[15] Trying to look rational, even at the price of some practical irrationality, may be the most rational thing to do.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

In the type of choices we have examined in this chapter, people's intuitions are generally weak. Having such weak intuitions is often a reliable sign that the decision at issue is not so important. After all, when it comes to dealing with the most pressing aspects of our ancestral environment, specific cognitive mechanisms are likely to have evolved and to provide us with strong intuitions. So, when our intuitions are weak, being guided by how easy it is to justify a particular decision is, in general, a simple and reasonable heuristic.

Our environments, however, have changed so much in the past millennia, centuries, and even decades that having weak intuitions is no longer such a reliable indication of the true importance of a decision. For most people, for instance, buying a car is an important decision. Much of their money goes into buying and maintaining a car; much of their time goes into using it; and their life may depend on its safety features. There are no evolved mechanisms for choosing cars in the way we have dedicated mechanisms aimed at selecting safe foods or reliable friends. As a result, intuitions give only weak and limited guidance. Does this mean that looking for an easily justifiable choice in these evolutionarily novel situations—choosing a car that is popular and well-reviewed, for instance—is unreasonably risky? Not really. In most cases, the decisions that are the easiest to justify in the eyes of others and hence that are the most likely to contribute to our reputation are also the best decisions to achieve our goals.

When the reasons that are recognized as good in a given community are objectively good reasons, people guided by reputational concern may still arrive at true beliefs and effective decisions. But this is not always the case—far from it. Throughout the centuries, smart physicians felt justified in making decisions that cost patients their lives. A misguided understanding of physiology such as Galen's theory of humors created a mismatch between decisions easy to justify—say, bleeding the patient to restore the balance between humors—and the fact that the condition of some patients deteriorated after being bled, a fact that must have given pause to some of these physicians. Still, if they were eager to maintain their reputation, they were better off bleeding their patients, and anyhow, there was no clear alternative. By contrast, today's doctors, relying on vastly improved, evidence-based medical knowledge, may make decisions guided in good part by the

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

sense of what the medical community would approve and, in so doing, preserve both their reputation and the health of their patients.

When Justification and Argumentation Diverge

The message of this chapter might seem bleak. Reason improves our social standing rather than leading us to intrinsically better decisions. And even when it leads us to better decisions, it's mostly because we happen to be in a community that favors the right type of decisions on the issue. This, however, cannot be the whole picture. Justifications in terms of reasons do indeed involve deference to common wisdom or to experts. What implicitly justifies this deference, however, is the presumption that the community or the experts are better at producing good reasons. But there is a potential for tension between the lazy justification provided by socially recognized "good reasons" and an individual effort to better understand and evaluate these reasons, to acquire some expertise oneself.

In the beginning of the nineteenth century, for instance, the doctor Joseph Victor Broussais was the most respected medical authority in Paris. He insisted that all fevers are caused by inflammation and should be treated by bloodletting. The younger doctor Pierre-Charles-Alexandre Louis didn't really doubt the efficacy of bloodletting, but he wanted to evaluate it precisely. For this, he compared two groups of patients that had been bled for pneumonia and discovered that, contrary to his expectations, those who had been bled early in their illness had died in greater numbers than those who had been bled late, showing that not only had bloodletting not cured them, it had worsened their condition. Louis had now compelling evidence and arguments, if not against bloodletting in general, at least against the systematic usage recommended by Broussais. Louis's pioneering work in evidence-based medicine played a crucial role in the progressive abandonment of bloodletting as a major medical procedure in the nineteenth century. In criticizing the overextended practice of bloodletting, Louis was taking immediate reputational risks, but precisely because he had good arguments to do so, in the end his ideas prevailed and his reputation grew.[16]

It would be nice to think that, when there is a conflict between the goal of having good reasons in the eyes of others and that of having demonstrably

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

good reasons, argumentative strength trumps ease of immediate justification and that the best reasons ultimately win. Well, things are somewhat more complicated.

Consider the following scenario:

> You have a ticket to a basketball game in a city sixty miles from your home. The day of the game there is a major snowstorm, and the roads are very bad. Are you more likely to go to the game if:
>
> a. You paid $35 for the ticket.
> b. You got the ticket for free.
> c. Equally likely.[17]

Most people answer that they would be more likely to face the snowstorm if they had bought the ticket than if they had got it for free. According to psychologists Hal Arkes and Peter Ayton, this decision is based on a reason such as "wasting is bad."[18] Most people would understand better why someone would brave a snowstorm for a ticket they bought than for one they got for free; they might even disapprove of somebody who bought the ticket and didn't go.

Economists call the price of the ticket in this situation a "sunk cost." The money has already been spent and cannot be retrieved. It is as good as sunk. Decisions are about the future, which can be altered, not about the past, which cannot. The only question that really matters, then, is: Would you be better off now facing the snowstorm to get to the game, or doing something else? If you would be better off doing something else, then undertaking an unpleasant and potentially dangerous drive simply makes you worse off. People who accept this argument should, it seems, answer that whether they bought the ticket or got it for free would not affect their decision to go or not to go to the game.

The argument against this so-called sunk-cost fallacy, on the other hand, is clear. It even has the backing of many philosophers and economists. If, however, you are convinced by the argument and you decide, say, to stay at home in spite of having paid for the ticket, you might be ill-judged by people who are not aware of this argument, and you might not have the opportunity to explain the reasons for your choice. So you might, ironically, be seen as making

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

an irrational decision when in fact you made a decision based on a sound reason. To the extent that you care what these people think of you, their judgment will pull toward making the socially acceptable decision.[19] (Mind you, if you are a student in economics and make a decision based on a sunk cost not to be ill-judged by your family, you may end up being deemed incompetent by your fellow economics students.)

But why should the sunk-cost fallacy be common and indeed be seen not as fallacious but as the right thing to do? Here is a speculative answer. One of the qualities people look for in friends, partners, or collaborators is that they should be dependable. Some degree of stubbornness in carrying through any decision made, in pursuing any course of action undertaken, even when the expectation of benefits is revised down, gives to others evidence that one can be relied upon. People who persevere in their undertakings even when it might not be optimally rational from their individual point of view may, in doing so, strengthen their reputation of reliability. It may be rational, then, at least in some cases, not just to knowingly commit the sunk-cost fallacy in order to signal to others that one can be counted upon but also to have a better opinion of people who commit the fallacy than of people who don't.

Attending to the interactional functions of reason not only makes better sense of it but also shows its limit. Justificatory and argumentative reasons are fundamental tools in human interaction, but which type of reason trumps the other when they diverge may depend not only on the quality of the reasons involved but also on the social, and in particular reputational, benefits involved. The reason module cannot pretend to the commandeering position that the classical approach assigned to capital *R* Reason.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

15

The Bright Side of Reason

At the beginning of the movie *12 Angry Men* (spoilers ahead), a youngster stands accused of stabbing his father to death. His life is in a precarious position: in the jury room, the arguments for conviction are piling up. One witness has seen the boy do it; another heard the fight and saw the accused flee out of the apartment; the boy's alibi doesn't hold water; he has a motive, and a long record of violence. Group polarization is lurking, ready to convince the jurors that the boy should be sent to the electric chair. But one juror is less confident than the others. While this juror is not convinced of the defendant's innocence, he's not quite sure of his guilt, either. He "just wants to talk." When urged to provide arguments, he starts with a weak one: the evidence against the boy is too good; it's suspiciously good. Unsurprisingly, this doesn't sway any of the other jurors. From then on, however, this juror does a better job at poking holes in the prosecution's case.

He unearths inconsistencies in the incriminating arguments. One witness claims to have seen the murder from the other side of the street, through the windows of a passing train of the elevated subway. Another witness claims to have heard the boy threaten his father—"I'll kill him!"—and then a body fall a few seconds later. But how could the second witness have heard anything amid the deafening sound of the train?

More inconsistencies emerge in other jurors' arguments. The boy's fingerprints cannot be found on the knife. Not a problem for juror four: the boy is a cold-blooded murderer who wiped the knife, still tainted with his father's blood. The defendant was caught by the police walking home three

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

hours after the crime. Why come back to the crime scene? Juror four has an answer: "He ran out in a state of panic after he killed his father, and then, when he finally calmed down, he realized that he had left the knife there." But then, points out the more skeptical juror, how do you square that with the fact that "he was calm enough to see to it that there were no fingerprints on the knife?"

Inconsistencies can take the form of double standards. The boy is from the slums, and juror ten knows they're "born liars"; "you can't believe a word they say." Yet he has no problems accepting the testimony of the witness who claims she saw the boy commit the murder, even though "she's one of them too," as our juror points out.

Sometimes the inconsistency surfaces so blatantly that it doesn't even need pointing out. Juror three has vehemently defended the guilty verdict, relying in large part on the testimony of the man who said he heard the fight and saw the kid leave the apartment right after. To see this, however, the witness claims that he got up and crossed the length of the whole building in fifteen seconds—something impossible for this old man with a limp. But that is not an issue for our juror three: the witness must have been mistaken in his estimate—after all, "he was an old man, half the time he was confused. How could he be positive about anything?"

These inconsistencies slowly sway most of the jurors toward reasonable doubt, but not all of them. Juror ten remains impervious to rational considerations. In the end, they must shame him to make him relent. Still, most of the work is done by argumentation. It is argumentation that allows the holes in the prosecution's case to surface. It is argumentation that highlights double standards. It is argumentation that lays bare inconsistencies. It is argumentation that raises doubt in the jurors' minds. In *12 Angry Men,* argumentation saves a boy's life.[1]

Argumentation Is Underrated

In Chapters 11 through 14, we have emphasized the "bad" sides of reason, those that have given rise to an enigma in the first place: reason is biased; reason is lazy; reason makes us believe crazy ideas and do stupid things. If we

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

put reason in an evolutionary and interactivist perspective, these traits make sense: a myside bias is useful to convince others; laziness is cost-effective in a back-and-forth; reason may lead to crazy ideas when it is used outside of a proper argumentative context. We have also repeatedly stressed that all of this is for the best—in the right context, these features of reason should turn into efficient ways to divide cognitive labor.

Crucially, this defense of argumentative reasoning depends on how people evaluate others' reasons: they have to be sensitive to what the philosopher Jürgen Habermas has called the "forceless force of the better argument."[2] They have to be able to reject weak arguments while accepting strong enough ones, even if that means completely changing their minds. In Chapter 12, we presented evidence that people are good at evaluating others' arguments, but these experiments still targeted solitary reasoners provided with a single argument to evaluate—a far cry from the back-and-forth of an argumentative exchange.

It is now time to make good on our promise and to show that in the right interactive context, reason works. It allows people to change each other's minds so they end up endorsing better beliefs and making better decisions.

For more than twenty years, Dave Moshman, psychologist and educational researcher, had asked his students to solve the Wason four-card selection task (the one from Figure 17) individually and then in small groups. While individual performance was its usual low—around 15 percent correct—something extraordinary was happening in the course of group discussions. More than half of the groups were getting it right. When Moshman teamed with Molly Geil to conduct a controlled version of this informal experiment, groups reached 80 percent of correct answers.

It may be difficult for someone who hasn't read article after article showing pitiful performance on the Wason four-card selection task to realize just how staggering this result is. No sample of participants had ever reached anywhere close to 80 percent correct answers on the standard version of the task. Students at the best American universities barely reach 20 or 25 percent of correct answers when solving the task on their own.[3] Participants paid to get it right still fail abysmally.[4]

What Moshman and Geil have achieved is the equivalent of getting sprinters to run a 100-meter race in five seconds by making them run together.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

You'd think that such an extraordinary result would get the attention of psychologists. Not at all: it went completely neglected. Perhaps no one really knew what to do with it. The only researchers who paid attention to Moshman and Geil's result were those whose theories were compromised. They asked for replications[5]—not unfairly, given the suspicious tendency of many psychology experiments not to replicate.[6] While not always as dramatic as in the original experiment, the improved performance with group discussion has proven very robust.[7] It also works very well with other tasks, such as the Paul and Linda problem we introduced in Chapter 12.[8] Try the experiment with friends, colleagues, or students—it works unfailingly.

Skeptical researchers also suggested that argumentation had little to do with the improvement in performance. Rather than paying attention to the content of each other's arguments, group members, they suggested, rely on superficial attributes to decide which answer to adopt. Perhaps people simply follow the most confident group member.[9] This alternative explanation makes some sense: confidence can be an important determinant of the outcome of group discussion, for better or worse.[10]

This lower-level interpretation, however, offers a very poor description of what happens when groups discuss a reasoning task. Looking at the transcripts, it is apparent that those whose views prevail are not just saying "I know that for a fact" with a confident tone. They put forward one argument after the other.[11] We also know that a single participant with the correct answer can convince a group that unanimously embraces the wrong answer, even if she is initially less confident than the other group members.[12]

How does the exchange of arguments fare when it's impossible to demonstrate, in the strict logical sense, that a given answer is correct and that the others are mistaken? When argumentation lacks such demonstrative force, other factors become more relevant in convincing people or in evaluating claims, such as who seems more competent, or how many people support a given opinion. Still, even for problems that do not have a single definite solution, group performance is generally above that of the average group member. In some cases it is even superior to that of the best individual in the group.[13] Incidentally, even when groups fail to surpass the answer of the best individual performer, one is still better off going with the answer of the group unless there is a clear way to tell who is the best performer in the first place.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

When Groups Work and When They Don't

Skepticism toward group efficiency is not entirely misplaced. When argumentation is not involved, group performance is disappointing. A hundred years ago, the agronomical researcher Maximilien Ringelmann noticed a weird pattern: tractors, horses, and humans seemed to be less efficient when performing a task jointly.[14] For instance, people pushed less hard to move a cart when they were doing it together.

Since the decrease in performance held for machines and animals as well as humans, Ringelmann assigned most of the blame to coordination problems: the strength is not applied simultaneously, which decreases the total strength exerted at any given time. However, observing prisoners powering a flour mill, he also noted that motivation could be an important factor for humans: "the result was mediocre because after only a little while, each man, trusting in his neighbor to furnish the desired effort, contented himself by merely following the movement of the crank, and sometimes even let himself be carried along by it."[15] Several decades later, social psychologists would show that such motivational factors are often the main culprits for group underperformance, labeling this phenomenon "social loafing."[16]

Groups can have disappointing performance not only when pooling physical force but also on a variety of cognitive problems. Brainstorming is a typical example. By and large, group brainstorming doesn't work. In a typical brainstorming session, participants are told not to voice their criticisms, so that they feel free to suggest even wild ideas. This doesn't work: a brainstorming group typically generates fewer and worse ideas than if the ideas of each individual working in isolation had been gathered.[17] By contrast, telling people that "most studies suggest that you should debate and even criticize each other's ideas" allows them to produce more ideas.[18]

That group performance should be disappointing in many domains only makes the successes of argumentation even more remarkable. When people argue, even about seemingly dull mathematical or logical tasks, there is no social loafing or cognitive disruption. Instead, their motivation is increased by the dialogical context. They respond to each other's arguments and build on them. Many great thinkers have noted the importance of a lively debate to fuel their intellect. Here is Montaigne:

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

> The study of books is a languishing and feeble motion that heats not, whereas conversation teaches and exercises at once. If I converse with a strong mind and a rough disputant, he presses upon my flanks, and pricks me right and left; his imaginations stir up mine; jealousy, glory, and contention, stimulate and raise me up to something above myself; and acquiescence is a quality altogether tedious in discourse.[19]

For a wide variety of tasks, argumentation allows people to reach better answers. The results reviewed so far, however, stem from laboratory experiments conducted in a highly controlled setting with participants who have not met before and will not see each other after the experiment. In the real world, things are different. Problems can be tremendously difficult and may not have a complete and satisfactory solution. Scientists look for the principles that govern the universe. Politicians try to get laws passed in deeply divided and confrontational parliaments. Judges search for a way to give due respect to legitimate but conflicting interests. In these situations, personal biases and affinities interfere or even take precedence. Strongly held convictions and values are attacked and staunchly defended. Does argumentation still have a positive role to play in managing more complex problems and overcoming emotional convictions?

How to Make Better Predictions

Prediction is hard, especially about the future. (Ironically this common aphorism also illustrates the difficulty of learning about the past, since it has been attributed, in one form or another, to everyone and their cousin, from Confucius to Yogi Berra.)[20] Phillip Tetlock, the expert of expert political judgment, wanted to find out just how hard it is to make good predictions in politics.[21] In the late 1980s he recruited 300 political experts, many with PhDs and years of experience, and asked them to do their job: make predictions about political events. Fast-forward fifteen years. The predictions are compared to the actual outcomes. How do the experts perform? Very poorly. They barely beat the proverbial dart-throwing chimp—random answers—and they are easily topped by simple statistical extrapolations from existing data.[22] Prediction *is* hard.[23]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

In a way, the main problem faced by the experts isn't so much that they weren't accurate but that they weren't aware that they weren't accurate. The world is a complicated place, and even experts face stringent cognitive limitations on the amount of information they can acquire and process. But that didn't stop them from making extreme forecasts: the experts were often saying that a given event was nearly certain to happen or not to happen.[24] Experts were much too confident in the power of their pet theories to predict the future.

In line with the experiments on polarization and overconfidence described in Chapter 14, Tetlock found that reasoning biases were responsible for these extreme and, more often than not, mistaken predictions. He observed that when making their predictions, experts have "difficulty taking other points of view seriously"[25] and that their "one-sided justifications are pumping up overconfidence."[26]

Reason also creates distortions in the way experts revise their beliefs. When an event happens as predicted by their favored theory, the experts grow more confident. But when things don't happen as they were expected to, our political experts turn into expert excuse finders. The war that failed to erupt is just about to be declared. A small accident of history prevented their predictions from coming true. Politics is too complicated anyway . . . [27] Some experts were so skilled at finding excuses that they became even more convinced that their theories were right after the events had proven them wrong.[28] Would the experts have been better off if they had been able to talk things over?

Predictions may have never mattered more to our species than during the Cold War, when the risk of an all-out atomic war was barely prevented by an "equilibrium of terror." The U.S. Air Force was one of the actors looking for better forecasts about the effects of nuclear war, and for this it turned to the RAND Corporation. Averaging the opinions of several experts offers a simple and efficient solution to improve on their forecasts. Yet two of RAND's researchers, Norman Dalkey and Olaf Helmer, thought they could do better by giving each expert information about what the other experts answered—such as the average answer, for instance. Experts made new predictions based on this information; the predictions were averaged and again provided to all the participants, who got to make another prediction; and so forth for a few rounds.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

This reiterated averaging technique, known as Delphi, was first used to figure out how many bombs would the Russians have to drop on U.S. industrial targets to reduce their output by three-quarters.[29] Fortunately, Dalkey and Helmer never found out if their method had yielded accurate forecasts in this specific case, but since the 1950s, many studies have shown that it can improve a variety of predictions, from defense issues to medical diagnoses.[30]

The Delphi method offers several advantages over face-to-face discussions. In face-to-face discussions, providing the best forecast may be less important than pleasing a senior colleague or keeping with the consensual opinion. Face-to-face discussions also require getting a group of busy experts in one room at the same time, which is not always easy to arrange. Delphi's anonymous questionnaires solve both problems.

Yet if the argumentative theory is right, the original form of Delphi is missing out on a major way of improving predictions: the exchange of reasons. If Zoe believes Italy has an 80 percent chance of winning the next soccer world cup, and Michael tells her he thinks it's only 20 percent, what should Zoe do? Balance the two opinions and adjust the odds to 50/50? On average, people put in this situation only go part of the way toward the other opinion. Zoe could settle on 60 percent, for instance. After all, she knows why she thinks Italy should have good odds, but she doesn't know the reasons for Michael's opinion.[31] If she knew why Michael is giving Italy low odds, she might be more inclined to change her mind.

In the Delphi method, instead of receiving only an average of others' forecasts, the experts can also be given the reasons for others' forecasts. Gene Rowe and George Wright looked at the difference this makes to the predictions.[32] As a matter of fact, the reasons did not make the experts change their minds more often. Indeed, they were more likely to cling to their initial opinion than when provided only with averages. But reasoning was not failing; it was merely revealing its discriminating power.

Not all reasons are good reasons. If Michael tells Zoe he thinks Italy has a 20 percent chance to win based on the predictions of an octopus,[33] she would be crazy to update her estimate at all. But if Michael tells Zoe he has insider information about the failing health of Italy's key forward, she might simply adopt Michael's odds. This is exactly what the participants in Rowe and Wright's experiment were doing. They were not changing their minds more,

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

but they were doing it more discriminately. They were changing their minds when they should, in the direction they should. And they were making better predictions.[34]

Providing a one-time summary of reasons is a good step forward, but it falls short of making the best of reasoning. Reasoning thrives in the back-and-forth of conversation, when people can exchange arguments and counterarguments. Online communication enables groups of experts to exchange arguments at a distance, opening up prospects for even better forecasts.

Twenty years after his original study of expert political judgment, Phillip Tetlock, together with Barbara Mellers and other colleagues, launched an even more ambitious experiment.[35] Over 1,300 participants were recruited and asked to make geopolitical predictions. Participants from a first group worked alone, kept in the dark about other participants' forecasts so as to maintain the independence of their judgments. Participants assigned to the second group also worked alone, but as in the early versions of the Delphi method, they were provided with statistical information about others' forecasts.

The third group was divided into teams of about twenty people who were allowed to discuss the forecasts together, online. Nearly all of their predictions proved more accurate than those of the independent forecasters, and they also beat the second group on two-thirds of the forecasts. Exchanging arguments had allowed them to produce significantly better predictions.

In Defense of Juries

So far we have seen that participants provide better answers to a variety of tasks when they are allowed to discuss them in small groups. We have seen that groups of experts can also take advantage of argumentation to improve their forecasts. But this chapter opened in a jury room—albeit a fictitious one—not a well-controlled experiment or an online chat among experts. The jury room is a typical face-to-face situation: tempers may flare, the search for consensus become paramount, errors never be dispelled. Jurors are the archetypal *non*experts, with only a tenuous understanding of the law. In *12 Angry Men,* the discussion is filled with prejudice, and many other biases cloud jurors' decisions.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

All too often, jurors form the wrong intuition regarding the verdict after hearing the evidence. With its myside bias, solitary reasoning is unlikely to help correct this initial intuition. Deliberation could turn into group polarization, amplifying rather than correcting shared prejudices. In light of these limitations, *12 Angry Men* seems optimistic indeed, an edifying and wishful ode to the power of argumentation. Perhaps we should follow the advice of Cass Sunstein and his colleagues: take some decisions out of juries' hands and give them to "specialists in the subject matter."[36]

Before damning juries, we should consider the alternative: judges and experts. Sir Edward Coke was undoubtedly such an expert. This English jurist of the late sixteenth and early seventeenth centuries was "possibly the most learned common lawyer of all time."[37] He was able to base his opinions "on innumerable medieval texts, most of them manuscript rolls, which he had perused with indefatigable zeal." All this perusing, however, was not antiquarian fancy. Coke "was clearly hoping to find precedents that would suit his legal and political convictions," and he "sometimes misinterpreted precedents to support his case." It may have been Coke that Sir William Blackstone had in mind when he warned, in his hugely influential *Commentaries on the Laws of England* of 1766, that a judge's knowledge and intelligence are no guarantee of fair opinion: "in settling and adjusting a question of fact, when entrusted to any single magistrate, partiality and injustice have an ample field to range in; either by boldly asserting that to be proved which is not so, or more artfully by suppressing some circumstances, stretching and warping others, and distinguishing away the remainder."[38]

Readers should not be surprised that judges, however competent, have a myside bias, using their erudition to defend preconceived opinions rather than arrive at an impartial verdict. Jurors are obviously not exempt from the myside bias, but, as Blackstone realized, deliberation has the potential to compensate for each juror's biases. Having berated judges, he continued: "Here therefore a competent number of sensible and upright jurymen . . . will be found the best investigators of truth, and the surest guardians of public justice." Two centuries after Blackstone, Justice Harry Blackmun of the U.S. Supreme Court would defend juries in similar terms: "the counterbalancing of various biases is critical to the accurate application of the common sense of the community to the facts of any given case."[39]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Blackstone may have been right in his pessimistic assessment of judges,[40] but was he also right that jury deliberation could balance out jurors' biases? This is not an easy question to answer. We do not have access to the arguments exchanged by jurors. At best, postdeliberation interviews can show that a view represented by a minority at the outset sometimes becomes the final verdict—we can tell at least that deliberation can change jurors' minds.[41] To know more about the effects of deliberation, we must rely on studies of mock juries.

In the early 1980s, Reid Hastie, Steven Penrod, and Nancy Pennington conducted a very important study of mock juries.[42] In order to make their experiment as realistic as possible, they recruited people who had been called for jury duty and showed them a three-hour video reenactment of a real trial before sending them in to deliberate. In this trial, the defendant stood accused of having stabbed a man to death during a fistfight that escalated. While the killing was well established, the verdict could plausibly range from not guilty—if the defendant were found to have acted in self-defense—to first-degree murder—if the defendant were found to have premeditated his crime. While it is impossible to know what the correct answer is for sure, according to the opinion of many legal experts, the appropriate verdict was second-degree murder; it was also the verdict delivered in the real trial that inspired the study.

Right after the jurors in the experiment had seen the three-hour video, they had to say which verdict they favored. The most common answer was manslaughter. Only a quarter of the jurors favored second-degree murder. In other words, most jurors initially got it wrong. Most juries, however, reached the best verdict. Even though the verdict of second-degree murder must have been defended by only a few jurors in some groups, these jurors often managed to convince the whole jury.[43] Deliberation had allowed many juries to reach a better verdict. Just like Blackstone expected, deliberation allowed jurors to counterbalance their respective biases.

Indeed, this is exactly what Phoebe Ellsworth, a scholar of law and psychology, observed while replicating the experiment of Hastie and his colleagues:

> Individual jurors tended to focus on testimony that favored their initial verdict preference: Testimony about the previous confrontation between

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

> the two men was generally raised by jurors who favored a murder verdict, whereas testimony that the victim punched the defendant immediately before the killing was generally raised by jurors who favored manslaughter or self-defense. This tendency is not a weakness, but rather a benefit of the deliberation process—the opportunity it affords for comparing several different interpretations of the events along with the supporting factual evidence.[44]

12 Angry Men is not that fanciful. As in the movie, individual jurors in real trials may be biased, they can make mistakes, and they certainly defend dubious interpretations. Yet deliberation makes jurors review the evidence more thoroughly and more objectively, compensating for individual biases and allowing juries to reach better verdicts. Ellsworth concluded her review of the movie with these optimistic words: "*12 Angry Men* is an Ideal, but it is an achievable Ideal."[45]

Argumentation Works

The interactionist approach to reason predicts that people should be good at evaluating others' reasons, rejecting weak ones and changing their mind when the reasons are good enough. Although this might seem like a trivial prediction, it runs against a general pessimism regarding the power of argumentation. For instance, people asked to estimate how easily participants would solve a logical task either on their own or in small groups don't think that groups would do much better than individuals.[46] Even psychologists of reasoning—who should know better—underestimate how well groups perform.

The results reviewed here belie this pessimistic but common view of argumentation. Again and again, we see people changing their minds when confronted with good arguments. Whether they solve logical tasks or look for new solutions to open-ended problems, whether they are experts or laypeople, whether they reflect on geopolitics or ponder what verdict to deliver, people reach better conclusions after debating the issue with their peers.

In Chapters 16 through 18, we will discover still more settings in which argumentation allows good ideas to spread and groups to outperform

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

individuals, showing that Thomas Jefferson was not unduly optimistic when he wrote that

> Truth is great and will prevail if left to herself, that she is the proper and sufficient antagonist to error, and has nothing to fear from the conflict, unless by human interposition disarmed of her natural weapons, free argument and debate, errors ceasing to be dangerous when it is permitted freely to contradict them.[47]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:16:57 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.