III

Rethinking Reason

What is reason? How does it work? What is it for? How could it evolve? In Chapters 7 through 10, we develop a novel interactionist approach that answers these questions. Reason, we argue, is a mechanism of intuitive inferences about reasons in which logic plays at best a marginal role. Humans use reasons to justify themselves and to convince others, two activities that play an essential role in their cooperation and communication. Just as echolocation evolved as an adaptation to the ecological niche inhabited by bats, reason evolved as an adaptation to a very special ecological niche, a niche that humans built and maintain for themselves with their intense social relationships, powerful languages, and rich culture.

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 7:17:40 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 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

7

How We Use Reasons

Humans appeal to reasons not just in reasoning but also in explaining and justifying themselves.[1] Nevertheless, explanation and justification on the one hand and reasoning on the other have been studied independently of one another, as if two quite different psychological mechanisms were involved. We believe that the difference is less one of mechanism than of function, and we want to articulate, in this third part of the book, an integrated approach to the psychology of reasons.

Why do you think this? Why did you do that? We answer such questions by giving reasons, as if it went without saying that reasons guide our thoughts and actions and hence explain them. These reasons are open to evaluation: they may be good or bad. Good reasons justify the thoughts or actions that they explain. This picture of the role of reasons in explanation and in justification may seem self-evident. It is based, however, on a convenient fiction: most reasons are after-the-fact rationalizations. Still, this fictional use of reasons plays a central role in human interactions, from the most trivial to the most dramatic.

The Commonsense Picture

Start with the dramatic: in the middle of the night of November 2, 2013, Theodore Wafer, a middle-aged man living in Dearborn Heights, Michigan, a white suburb of Detroit, was awakened by loud banging on his front door. Earlier that night, Renisha McBride, a young African American woman, had crashed her car, walked out, and wandered for hours in a state of confusion. She ended

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

up at Wafer's door, probably looking for help. Wafer thought his house was being attacked, took his shotgun, opened the door, and fired, killing McBride. At the trial, he explained his action by saying that he had been afraid for his life and had shot in what he believed was self-defense against several attackers. An unreasonable fear, the prosecution argued. In any case, even if the fear had been reasonable, the reasonable response would have been not to open the door but to lock himself in and call the police. Wafer was found guilty of second-degree murder.

Why did Wafer act the way he did? No doubt the reasons he gave at the trial were the best he could muster. We will never know for sure, and it may well be that he himself didn't know what exactly had gone on in his mind at the time. But say we accept that he acted for the reasons he himself invoked: he felt under attack and wanted to defend himself. Even so, most of us would agree with the prosecution and the jury that he didn't have good enough reasons to believe he was in great danger or to act the way he did.

As the case of Wafer illustrates, some reasons may seem good enough to explain but not good enough to justify. We can accept an explanation and, at the same time, be critical of the reasons it invokes. For the purpose of explanation, it is enough that these reasons should have seemed adequate to the person we are trying to understand. On the other hand, to judge that what the person thinks or does is justified, the same reasons must seem adequate *to us*.

Much less dramatically, we all invoke minor or even minute reasons at every turn in daily social interactions. Rob, for instance, asks Ji-Eun, "Do you want a milkshake?" and she answers, "Thank you, but, you know, most of us Koreans are lactose intolerant." Why doesn't Ji-Eun just say no? Why does she bother to give a reason for her refusal? By mentioning a reason for declining Rob's offer of a milkshake, Ji-Eun suggests that she is appreciative of the offer and might have accepted it otherwise. Rob is unlikely to question the reason invoked by Ji-Eun. What should matter to him is that by giving a reason rather than just saying, "No!" Ji-Eun has been considerate toward him. In such ordinary interactions, our giving reasons manifests the kind of consideration others can expect of us and we might expect of them.

How must people understand reasons to be able to use them as they do in their thinking and interactions? To answer this question, it would be of lim-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

ited use to look at the way people understand the word "reason" itself (in the sense of a reason for something). This understanding varies across people, across circumstances, and, obviously, across cultures. The English word "reason" in that sense doesn't have straightforward translations in all languages. This fuzziness and this variability in the way people think and talk about reasons in general are worth studying in their own right. What we are investigating here, though, are not folk notions or folk theories of reasons but the way people make use of particular reasons and do so whether or not they categorize them as "reasons."

To make better sense of the way reasons are produced and used, it will be helpful to make a distinction between objective and psychological reasons. Such a distinction (with subtle differences and various terminological choices) is common in the rich philosophical literature on reasons. For our present purpose we will, when useful, draw inspiration from this literature without getting involved in its controversies.[2]

An *objective reason,* as we will use the phrase, is a fact that objectively supports some conclusion. This conclusion may be descriptive (about what is the case) or practical (about what is desirable). For instance, the fact that today is Friday is an objective reason to conclude that tomorrow will be a Saturday; the fact that the plums are ripe is an objective reason to pick them from the tree without waiting.

Facts, as we are using the term, are true propositions, abstract objects without causal powers. It is not, strictly speaking, the fact that some plums are ripe that causes them to fall from the tree; it is their ripeness itself or, to be more precise, it is the chemical and physical changes that define ripeness. Objective reasons, being facts, are themselves without causal powers. From our empirical point of view, it doesn't matter whether objective reasons exist in the world independently of human interests and what well-defined criterion of objectivity they meet, if any. If we talk about objective reasons at all, it is because they are represented both in what people think and in what they say and, unlike facts, representations of facts do have causal powers.

A *psychological reason* is a mental representation of an objective reason. Like all mental representations, it may be a misrepresentation; that is, it may represent as an objective reason for a conclusion a false proposition rather than a fact, or it may represent a genuine fact as supporting a conclusion it actually

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

does not support. Still, in our cognitive and evolutionary perspective, we assume that in general, people tend to represent as objective reasons for a given conclusion facts that do indeed support it: cognition is imperfect but not random. Psychological reasons are mental representations in the brain and, as such, play a causal role in people's lives. (When we use "reasons" without qualifying the term, we are talking about psychological reasons.)

It is generally thought that the main role psychological reasons play is to motivate and guide people's actions and beliefs (guidance being little more than a fine-grained form of motivation and motivation a coarse-grained form of guidance). We disagree. The main role of reasons is not to motivate or guide us in reaching conclusions but to explain and justify after the fact the conclusions we have reached.

The uses of reasons to explain and to justify are not just related; they are intertwined. To explain people's beliefs or decisions, psychological reasons must at least point in the direction of a conceivable justification, that is, of a good, objective reason. Wafer, for instance, could, perhaps not with sufficient good sense but not absurdly either, see his reasons as objectively justifying his shooting the person at the door. Had he, on the other hand, given as his reason for having fired a shot that Elvis Presley was dead and that therefore life was meaningless, we would see this not as a genuine reason-based explanation, not even a defective one, but as an admission (or a claim) of temporary insanity.

Similarly, when we invoke reasons to justify other people's thoughts or actions, the normal implication is that they had these reasons in mind and were guided by them. How could people be justified by reasons they knew nothing about? This seems to be a commonsensical constraint on justification, but is it really? Actually, there are interesting exceptions described in the philosophical literature as cases of "moral" or "epistemic luck."[3]

Imagine, for instance, what would have happened if the person Wafer shot and killed had turned out to be not a young woman looking for help but a dangerous criminal on the "Most Wanted" list. In that case, Wafer's action might well have been commended even though he would have acted without knowing that there happened to be good, objective reasons to do what he did—a typical case of moral luck. Reasons must be in people's minds to ex-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

plain their behavior, but in some cases at least, they need not be in their minds to make their behavior either blameworthy or praiseworthy.

Just as there are cases of moral luck—an action undertaken with inadequate reasons in mind but that turns out to have had, unbeknownst to the agent, an excellent objective justification—there are cases of epistemic luck—a belief held for inadequate reasons but that turns out to be justified all the same.

The Alexandrian scholar Eratosthenes, as we saw in Chapter 1, was the first person ever to measure the circumference of the earth. The result of his calculation, 252,000 stades, was just 1 percent shy of the modern measurement of 24,859 miles, or 40,008 kilometers. What makes this precision even more astonishing is that Eratosthenes had made two mistakes in his assumptions. It was essential to his calculation that the town of Syene should be right on the Tropic of Cancer and due south of Alexandria, and he was convinced that such was indeed the case. Actually, Syene is 1° north of the Tropic of Cancer and 3° east of Alexandria. So how could he come so close? By pure chance, the effects of these two errors practically canceled one another out. This bit of epistemic luck has not been used to downplay Eratosthenes's accomplishment. It is as if his overall method was so brilliant and his result so impressive that the inadequacy of some of his reasons was irrelevant.

Moral and epistemic luck are puzzling phenomena: they suggest that we may find people justified by reasons that actually didn't motivate or guide them and that therefore cannot explain their thoughts or actions. Could the commonsense picture of reasons not be as clear and coherent as it seems at first sight?

The Commonsense Picture Challenged

Contrary to the commonsense picture, much experimental evidence suggests that people quite often arrive at their beliefs and decisions with little or no attention to reasons. Reasons are used primarily not to guide oneself but to justify oneself in the eyes of others, and to evaluate the justifications of others (often critically). When we do produce reasons for guidance, most of the time it is to guide others rather than ourselves. While we would like others to be guided by the reasons we give them, we tend to think that we

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

ourselves are best guided by our own intuitions (which are based, we are sure, on good reasons, even if we cannot spell them out).

Whether or not it would be better to be guided by reasons, the fact is that in order to believe or decide something, we do not need to pay any attention to reasons. Purely intuitive inference, which generates so many of our beliefs and decisions, operates in a way that is opaque to us. You look at your friend Molly and somehow intuit that she is upset. What are your reasons for this intuition? Or you check what films are playing tonight at the Odeon: *Star Wars 12* and *Superman 8.* You decide to go and see *Superman 8.* What are your reasons for this choice? If asked, sure, you would produce reasons, but the fact is that at the moment of intuiting that Molly was upset or of choosing *Superman 8,* you were not consciously entertaining, let alone pondering, reasons. The opinion and the choice came to you intuitively.

Still, one might object, reasons may well have guided us unconsciously. Moreover, we are generally able to introspect and to become conscious of our unconscious reasons. But is this really what is happening? When we explain ourselves, do we really bring to consciousness reasons that have guided us unconsciously? Let us first look at some challenging evidence and, in the next section, propose an even more radical challenge.

The commonsense confidence in one's ability to know one's mind was, of course, undermined by the work of Sigmund Freud and its focus on what he called "the Unconscious." The existence of unconscious mental processes had been recognized long ago, by Ptolemy or Ibn Al-Haytham, but until Freud, these processes were seen as relatively peripheral. Mental life was regarded, for the most part, as typically conscious, or at least open to introspection. However, Freud made a compelling case that we are quite commonly mistaken about our real motivations. A century later, in a cognitive psychology perspective, the once radically challenging idea of the "Unconscious" seems outdated. Not some, but all mental processes, affective and cognitive, are now seen as largely or even wholly unconscious. The problem has become, if anything, to understand why and how there is something like consciousness at all.[4] Freud's challenge to the idea that we know our reasons has been, if anything, expanded.

In a very influential 1977 article, two American psychologists, Richard Nisbett and Timothy Wilson, reviewed a rich range of evidence showing that we

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

have little or no introspective access to our own mental processes and that our verbal reports of these processes are often confabulations.[5] Actually, they argued, the way we explain our own behavior isn't that different from the way we would explain that of others. To explain the behavior of others, we take into account what we know of them and of the situation, and we look for plausible causes (influenced by the type of causal accounts that are accepted in our culture). To know our own mind and to explain our own behavior, we do the same (drawing on richer but not radically different evidence). In his book *The Opacity of Mind,*[6] philosopher Peter Carruthers showed how much recent research had confirmed and enriched Nisbett and Wilson's approach (which he expands both empirically and philosophically).

Our inferences about others are often quite insightful; our inferences about ourselves needn't be worse. We may often succeed in identifying bits of information that did play a role in our beliefs and decisions. Where we are systematically mistaken is in assuming that we have direct introspective knowledge of our mental states and of the processes through which they are produced.

How much does the existence of pervasive unconscious processes to which we have no introspective access challenge our commonsense view of ourselves? The long-established fact that the operations of perception, memory, or motor control are inaccessible to consciousness isn't really the problem. Much more unsettling is the discovery that even in the case of seemingly conscious choices, our true motives may be unconscious and not even open to introspection; the reasons we give in good faith may, in many cases, be little more than rationalizations after the fact.

We have already encountered (in Chapter 2) a clear example of such rationalization from the psychology of reasoning. In the Wason four-card selection task, participants, before they even start reasoning, make an intuitive selection of cards. Their selection is typically correct in some versions of the task and incorrect in others, even though the problem is logically the same in all versions. Asked to explain their selection, participants have no problem providing reasons. When their choice happens to be correct, the reasons they come up with are sound. When their choice happens to be mistaken, the reasons they come up with are spurious. In both cases—sound and spurious reasons—these are demonstrably rationalizations after the fact. In neither case are participants

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

aware of the factors that, experimental evidence shows, actually drove their initial selection (and which are the same factors whether their answer is correct or mistaken). Still, such experimental findings, however robust, smack of the laboratory.

Fortunately, not all experimental research is disconnected from real-life concerns. The brutal murder of Kitty Genovese in New York, on March 13, 1964, with dozens of neighbors who heard at least some of her cries for help but didn't intervene, prompted social psychologists John Darley and Bibb Latané to study the conditions under which people are likely or unlikely to help.[7] They discovered that when there are more people in a position to be helpful, the probability that any of them will help may actually decrease. The presence of bystanders causes people to ignore a person's distress (a phenomenon Darley and Latané dubbed "the bystander effect"), but this is a causal factor of which people are typically unaware.

In one study (by Latané and Judith Rodin),[8] people were told that they would participate in a market study on games. Participants were individually welcomed at the door of the lab by a friendly assistant who took them to a room connected to her office, gave them a questionnaire to fill out, and went back to her office, where she could be heard shuffling paper, opening drawers, and so on. A while later, the participant heard her climb on a chair and then heard a loud crash and a scream, "Oh, my God, my foot . . . I . . . I . . . can't move it. Oh . . . my ankle! . . . I . . . can't get this . . . thing . . . off me."

In one condition, the participant was alone in the room when all this happened. In another condition, there was a man in the room who acted as if he were a participant too (but who was, in fact, a confederate of the experimenter). This man hardly reacted to the crash and the scream. He just shrugged and went on filling out the questionnaire. When real participants were on their own in the room, 70 percent of them intervened to help. When they were together with this apparently callous participant, only 7 percent of them did.

Immediately after all this happened, participants were interviewed about their reactions. Most of those who had taken steps to help said something like: "I wasn't quite sure what had happened; I thought I should at least find out." Most of those who didn't intervene reported having thought that whatever had happened was not too serious and that, moreover, other people working

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

in nearby offices would help if needed. They didn't feel particularly guilty or ill at ease. Had it been a real emergency, of course, they would have helped (or so they claimed).

When asked whether the presence of another person in the room had had an effect on their decision not to help, most were adamant that it had had no influence at all. Well, we know that they had been ten times more likely to intervene when they were alone in the room than when they were not. In other terms, the presence of that other person had a massive influence on their decision. Various factors can help explain this "bystander effect": when there are other people in a position to help, one's responsibility is diluted, the fact that other people are not reacting creates a risk of appearing silly if one does, and so on. What is relevant here is the forceful demonstration of how badly mistaken one can be about what moves one to act or not to act (expect more striking examples in Chapter 14).

What is happening? Do we form beliefs and make decisions for psychological reasons that are often unconscious, that we are not able to introspect, and that we reconstruct with a serious risk of mistake? Or is what generally happens even more at odds with the commonsense view of the role of reasons in our mental life?

Modules Don't Have Reasons

The evidence we have considered so far suggests that humans have limited knowledge of the reasons that guide them and are often mistaken about these reasons. We want to present an even more radical challenge to the commonsense picture. It is not that we commonly misidentify our true reasons. It is, rather, that we are mistaken in assuming that all our inferences are guided by reasons in the first place. Reasons, we want to argue, play a central role in the after-the-fact explanation and justification of our intuitions, not in the process of intuitive inference itself.

Of course, few philosophers or psychologists would deny the obvious fact that we often form beliefs and make decisions without being conscious of reasons for doing so. Still, they would argue that whether or not we are conscious of our reasons, we are guided by reasons all the same. It is just that these reasons are "implicit." This is what happens in intuitive inference. But what

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

are implicit reasons? How can they play their alleged guiding role without being consciously represented?

The word "implicit" is borrowed from the study of linguistic communication, where it has a relatively clear sense. On the other hand, when psychologists or philosophers talk of implicit reasons, they might mean either that these reasons are represented unconsciously or that they aren't represented at all (while somehow still being relevant). Often, the ambiguity is left unresolved, and talk of "implicit reasons" is little more than a way to endorse the commonsense view that people's thought and action must in some way be based on reasons without committing to any positive view of the psychological reality and actual role of such reasons.[9]

We believe that the explicit-implicit distinction is a clear and useful one only in the study of verbal communication and that the only clearly useful sense in which one may talk of "implicit reasons" is when reasons are implicitly communicated. When, for example, Ji-Eun answers Rob's offer of a milkshake by saying, "Thank you, but, you know, most of us Koreans are lactose intolerant," the reason she gives for her refusal is an implicit reason, in the linguistic sense of "implicit": it is not explicit—she doesn't say, for instance, that she herself is lactose intolerant—but it can and should be inferred from her utterance.

Still, one might maintain that psychological reasons can be conscious or unconscious, that at least some unconscious reasons can be made conscious, and that it makes sense to call unconscious reasons that can be made conscious "implicit reasons." But are there really implicit reasons in this sense? We doubt it.

Unconscious and intuitive inferences are carried out, we argued, by specialized modules. Modules take as input representations of particular facts and use specialized procedures to draw conclusions from them. When a module functions well and produces sound inferences, the facts represented in the input to the module do indeed objectively support the conclusion the module produces. This, however, is quite far from the claim that the representations of particular facts are unconscious reasons that guide the work of the module.

Here is why not. A fact—any fact—is an objective reason not for one conclusion but for an unbounded variety of conclusions. The fact, for instance, that today is Friday is an objective reason to conclude not only that tomorrow

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

will be a Saturday but also that the day after tomorrow will be a Sunday; that tonight begins the Jewish Shabbat; that, in many offices, employees are dressed more casually than on other working days of the week; and so on endlessly.

The same fact, moreover, may be a strong objective reason for one conclusion and a weak one for another. For instance, the fact that the plums are ripe is a strong reason to conclude that if they are not picked they will soon fall and a weaker reason to conclude that they are about to be picked. The same fact may even be an objective reason for two incompatible conclusions. For instance, the fact that it has been snowing may be a reason to stay at home and also a reason to go skiing.

It follows from all this that the mental representation of a mere fact is not by itself a psychological reason. The representation of a fact is a psychological reason only if this fact is represented as supporting some specific conclusion. I may know that you know that it has been snowing and not know whether this fact is for you a reason to stay home, a reason to go skiing, a reason for something else, or just a mere fact and not a reason for any particular conclusion at all. We cannot attribute reasons to others without knowing what their reasons are reasons for. Well, you might think, so what? Surely, if the representation of a fact (real or presumed) is used as a premise to derive a conclusion, then it is a psychological reason for this very conclusion. This, however, is mistaken. A belief used as a premise to derive a conclusion is not necessarily a psychological reason for this conclusion.

A long time ago, Ibn Al-Haytham argued that the mind performs unconscious inferences by going through the steps of a syllogism. If this were truly the case, then the representation of a particular fact would serve as the minor premise of such a syllogism, the regularity that justifies inferring the conclusion from this particular fact would serve as the major premise, and these two premises taken together could be seen as a psychological reason for the conclusion of the syllogism (each premise being a partial reason in the context of the other premise). This logicist understanding of all inferences, conscious or unconscious, is still commonly accepted, and this may be why it may seem self-evident that whatever information is used as an input (or a "premise") to an inference has to be a reason for the conclusion of this inference. The view, however, that unconscious inferences are produced by going through the steps of a syllogism or more generally through the steps of a logical derivation has

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

been completely undermined by modern research in comparative psychology and in psychology of perception (as we argued in Chapter 5).

Unconscious inferences, we argued, are produced by modules; these modules exploit the regular relationship that exists between the particular facts they use as input and the conclusion they produce as output, but they don't represent this relationship either as the major premise of a syllogism or in any other way. Representations of particular facts, we pointed out, are not by themselves psychological reasons for any particular conclusion. Modules, in any case, don't need reasons to guide them. They can use representations of facts as input without having to represent, either as a reason or in any other way, the relationship between these facts and the conclusions they derive from them. Modules don't need motivation or guidance to churn out their output.

Consider, as a first illustration, the case of a rudimentary inference that, even though it takes place in the nervous system, doesn't, properly speaking, take place in the mind. Perspiration is a well-developed mechanism in humans (and in horses).[10] When body temperature rises too much, the hypothalamus, a brain structure, triggers the production of sweat that cools the body. This is obviously something that happens in us and to us rather than something that we intend. Still, the perspiration mechanism performs a rudimentary practical inference: it takes as input information about body temperature provided by various neural detectors and, when appropriate, it yields as output instructions to the sweat glands. In doing so, it exploits the general fact that, above some temperature threshold, sweating is appropriate, but it does not *represent* this general fact. Information about current body temperature, which *is* represented in the module, isn't by itself a reason for anything in particular. The module clearly does its job without being guided by any psychological reason, and it doesn't need any such reason to perform its job.

As a second example, take desert ants. Once they have found some piece of food, they speed back to their nest in an almost straight line (as we saw in Chapter 3). Ants know where to go thanks to what Wehner calls a "navigational toolkit," a complex cognitive module with specialized submodules. The procedures used by the submodules (counting steps, taking into account angular changes of trajectory, and so on) each evolved to take advantage of a reliable regularity without, however, representing this regularity. On each foray the ants make outside the nest, relevant information inferred by these submod-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

ules is automatically integrated and contributes to determining the ants' return path. In explaining this process, there is no ground to assume that ants are guided by explicit reasons or that the modules involved are guided by implicit reasons.

Why do we bother, you may ask, to make the obvious point that human perspiration and desert ant orientation are not guided by reasons, either explicit or implicit? Because considerations that are relevant in these two cases extend quite naturally to less straightforward cases and, to begin with, to the case of inference in perception.

Remember the Adelson checkerboard illusion we talked about in Chapter 1 (see Figure 4)? Participants are asked which of two squares in the picture of a checkerboard is darker when, actually, both are exactly the same shade of gray. They see—here is the illusion—one of the two squares as much lighter than the other. They do so because they infer the relative darkness of a surface not just from the light it reflects to their eyes but also from the light they assume it receives. One of the two squares is depicted as being in the shadow of a cylinder and therefore as receiving less light than the other. Participants automatically compensate for this difference in light received, and see this square as lighter than it is.

In this textbook example of the role of inference in perception, no one would argue that we have conscious reasons to see one square as darker than the other. The module involved computes the relative darkness of each square as the ratio of the light the square reflects to the light it receives because it evolved to do so; in nonillusory cases, there are objective reasons why it should do so; these reasons, however, are not represented in the mechanism. The module is not guided by unconscious psychological reasons.

Some basic perceptual inferences, we hope you now agree, are not guided by psychological reasons. But does the argument extend to less automatic, more interpretive aspects of perception and to intuitive inference generally? When participants in the Latané and Rodin experiment heard various noises coming from the next room, they recognized some of these as noises of impact, others as words spoken in English. This recognition was, presumably, quite automatic: mentally represented reasons had no role to play in the process (even though, of course, it was an inferential process). Putting these perceptions together and inferring that the woman next door had fallen and hurt herself

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

involved more interpretation, understanding that she might need help even more so. Does this mean that the participants had to have explicit reasons to come to these further more interpretive conclusions or that the mechanisms involved had to have unconscious reasons?

Any socially competent person familiar with the kind of modern environment where the experiment was taking place would have developed a modularized capacity to make the relevant inferences. They would spontaneously recognize some noises not just as noises of impact but, using subtle acoustic features, as noises of a crash involving both hard pieces of furniture and a softer body such as a human body falling on a hard surface. We suggest that the inferential perception procedure that permits such recognition exploits correlations between features and probable causes of noises without representing these correlations, let alone representing them as reasons.

When the participants heard the words, "Oh, my God, my foot . . . I . . . I . . . can't move it. Oh . . . my ankle! . . . I . . . can't get this . . . thing . . . off me," they spontaneously understood "it" to refer to the speaker's foot, they understood "this . . . thing" to refer to some large solid object like a piece of office furniture that had fallen on the person's foot, and in so doing they were going well beyond what "it" and "this thing" linguistically encode. To perform these pragmatic inferences, participants used a modular comprehension procedure that exploits, without representing them, relevance-based regularities in verbal communication.[11] They homed in on one of many linguistically possible interpretations without representing, either consciously or unconsciously, reasons to do so.

Having understood in this intuitive manner that the woman next door had hurt herself and was communicating about her predicament, most participants would spontaneously conclude that she needed help. Again, we suggest that they could come to that conclusion without having to mentally represent (consciously or unconsciously) a reason for the inference. In fact, if someone had to mentally represent such a reason in order to realize that a person in pain and complaining might need help, this absence of spontaneous empathy might be taken as a symptom of impaired cognitive and social competence.

Still, there is no doubt that reasons did play a role in Latané and Rodin's experiment. The question is when reasons appeared in the course of events. After the crash in the next office, some participants stood up and went to see

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

whether they could help. Others remained seated. Was their decision either to try to help or to do nothing a spontaneous decision arrived at without considering reasons? We would suggest that unless they showed some hesitation, there is no strong ground to assume that they had been guided by a conscious or unconscious reason represented in their mind. Be that as it may, a few minutes later, when the experimenter asked them what had gone through their mind, they readily provided reasons that, they claimed, had motivated their action or inaction.

Some participants, especially those who had been helpful, may have come up with reasons just when they were asked. For them, finding reasons was effortless: in circumstances where it is obvious that someone may need you to help, wanting to help is an obviously good reason. Those who had remained seated and had done nothing had had more time to think about reasons they might invoke to justify themselves. They may also have felt a greater need to think of such a justification. Still, even for them, devising a plausible justification was not that difficult. In both cases, thinking about personal reasons for one's decision typically came after having made the decision itself.

If, as we suggest, the point of reasons isn't to guide the formation of beliefs and the making of decisions, then what are reasons for?

Reasons Are for Social Consumption

Whatever humans do is likely to contribute for better or worse to the way they are seen by others—in other words, to their reputation. These indirect reputational effects may turn out to be no less important than the direct goal of their action, whatever it is. Socially competent people are hardly ever indifferent to the way their behavior might be interpreted. By explaining and justifying themselves, people may defend or even improve their own reputation. By failing to do so, they may jeopardize it.

Thinking about good reasons for their actions is something that people often do proactively, anticipating that they may be called upon to explain or justify themselves. The minute you have engaged in a course of action that may have reputational costs—and sometimes even before, when you are merely considering it—a different mental mechanism may start working. Its function is to manage your reputation and for this, to provide an explanation that will

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

justify your behavior. Participants in the Latané and Rodin experiment who ignored the events taking place in the next room may have been aware that their passivity might be open to criticism. This may have prompted them to think of reasons they might provide to justify themselves.

The reputation management mechanism acts like a lawyer defending you, whatever you have done. Still, given the opportunity, a lawyer may advise a regular client against a course of action that would be hard or impossible to defend. What happens when the reputational mechanism cannot come up with a good narrative? Then the course of action considered (and possibly already undertaken) turns out to have costs that weren't initially taken into account in the decision process. In such a case, the failure of the reputational mechanism to produce an adequate narrative may have a feedback effect and cause the initial decision to be rescinded, or at least revised. It is possible, for instance, that a few of the participants in the Latané and Rodin experiment who had first dismissed the idea of helping the person in distress in the next room then realized that they would be hard put to justify their passivity; hesitated, that is, compared the ease with which they could justify their two possible responses; and chose to do something after all.

There is some fascinating experimental evidence that the search for reasons aimed at justification may, in fact, influence action (more about this in Chapter 14). Still, living up to the story you want to be able to tell about yourself isn't quite the same thing as telling a true story.

So, when people produce reasons to explain and justify their beliefs or actions, the narrative they come up with may be at odds with what really happened in their mind in three different ways, each interesting in its own right.

First, people commonly present themselves as having considered reasons and been guided by them in the process of reaching a belief or a decision that, in fact, was arrived at intuitively. The error we all make here is to falsely assume that we have direct knowledge of what happens in our mind when we draw intuitive inferences and that we are guided by reasons of which we are conscious or that we can easily introspect. Contrary to such an assumption, in coming to an intuitive decision or belief, we are not guided by reasons (not even "implicit" ones).

Even so, we often do correctly identify a piece of information that served as input to our inferential processes. When this happens, our mistake is just

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

to describe this piece of information as a personal reason. Participants who went to see if they could help the person in the next room had mentally represented the fact that that person was in pain and needed help; what they hadn't mentally represented at the moment was the higher-order consideration that the person needing help was a reason to go and help. This appropriateness of helping people in need (in certain circumstances at least) is so obvious that in socially competent people, it is exploited by a dedicated procedure, which works without representing it. For our social interactions, such a mistaken self-attribution of reasons doesn't matter at all. In fact, it is less an individual mistake than a socially encouraged use of commonsense psychology. From a scientific point of view, on the other hand, this should be recognized as a misrepresentation: acting spontaneously to help others need not be guided by conscious or unconscious reasons any more than smiling at someone who smiles at you.

Second, people may be mistaken about the information that served as input to their inferences. Participants who chose not to help were first and foremost influenced by the presence and passivity of the other person in the room, but they failed to acknowledge this crucial factor. In all likelihood, when they claimed that their decision had been motivated by the thought that nothing too serious had happened and that their help was not needed anyhow, they were just reporting thoughts that occurred to them only after they had made the intuitive decision to align their behavior to that of the other person in the room. In giving these reasons to the experimenter, they may have been sincere, even if self-deluded: people who think of themselves as nice and helpful and who realize they didn't help when they should have may be puzzled by their own behavior. The easy, too easy, solution to this puzzle is to assume that they must have had what looked at the time like good reasons to think that their help wasn't required.

Third, there are people who fail to find good enough reasons for what they are about to do and who, as a result, waver and change or at least readjust their course of action. In such cases, the personal reasons people invoke have truly played a causal role in their final decision. These reasons didn't directly shape the decision that was finally taken, but they inhibited the carrying out of the initial decision. Still, contrary to what these people may believe and claim, the true causal process didn't go from reasons to decision, but from

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

tentative decision to search for justification, failure of search, and revised decision aimed at a more easily justifiable course of action.

What we have said so far implies that there is no role for unconscious reasons in human cognition. If unconscious reasons have no role to play, then they probably don't exist, and if they don't exist, then implicit reasons don't exist either. When we attribute to ourselves an implicit reason, we are just interpreting our thoughts or actions in terms of newly constructed conscious reasons that we fictitiously represent as having been implicit before.

Does all this suggest that psychological reasons, whether "implicit" or conscious, have no reality whatsoever, that they are a pure construction? No, reasons are indeed constructed, but under two constraints that ensure that they have some degree of both psychological and social reality. The reasons we invoke for justification have to make psychological sense. Talk of reasons need not—in fact, we have argued, cannot—provide an accurate account of what happens in our minds, but it tends to highlight factors that did play a causal role. Reasons are typically constructed out of bits of psychological insight.

In order to fulfill their justificatory function, moreover, psychological reasons have to represent objective reasons socially recognized as good reasons. This social recognition may be biased by many social and cultural factors, but if it were a purely arbitrary affair, how would it hold any sway? For many beliefs and practices—and in particular for the use of reasons as justification or argument—genuine objectivity is, we maintain, a major factor in cultural success (a fact well explained by our account of the evaluation of reasons, as discussed in Chapter 13).

To attribute reasons to oneself or to others is not so much to formulate a hypothesis about how things are as to construct a tool for social action: justification, evaluation, or, we will see in Chapters 8 through 10, argumentation. But what makes reasons a useful tool?

When we give reasons for our actions, we not only justify ourselves, we also commit ourselves. In the first place, by invoking reasons, we take personal responsibility for our opinions and actions as described by us, that is, as attitudes and behavior that we had reasons to adopt. We thereby indicate that we expect others to either accept that we are entitled to think what we think and do what we do or be ready to challenge our reasons. When what we

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

thought or did is unlikely to be approved, by giving reasons, we may indicate a line of defense: we had, if not good reasons, at least reasons that *seemed* good at the time. A defense based on reasons typically allows us to accept responsibility while denying guilt.

By giving reasons, we also commit ourselves to a future line of thought and conduct. Invoking reasons as motivations of one's past views and actions expresses a recognition of the normative aptness of these reasons and a commitment to being guided by similar reasons in the future. For our audience, this commitment to accepting responsibility and to being guided in the future by the type of reasons we invoked to explain the past is much more relevant than the accuracy of our would-be introspections. This is why we all pay attention to the reasons of others and why we produce our own.

To put it in more sociological terms: Reasons are social constructs. They are constructed by distorting and simplifying our understanding of mental states and of their causal role and by injecting into it a strong dose of normativity. Invocations and evaluations of reasons are contributions to a negotiated record of individuals' ideas, actions, responsibilities, and commitments. This partly consensual, partly contested social record of who thinks what and who did what for which reasons plays a central role in guiding cooperative or antagonistic interactions, in influencing reputations, and in stabilizing social norms. Reasons are primarily for social consumption.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

8

Could Reason Be a Module?

On a sunny day, an elder from the Dorzé tribe in southern Ethiopia was exhorting a group of idle young men: they shouldn't smoke cigarettes—it was against their religion! The young men pointed out that the young French anthropologist (Dan) who had been living among the Dorzé for the past few months was sitting nearby, doing what? Smoking cigarettes! The elder turned to the foreigner and demanded an explanation: "Why do you smoke?" The anthropologist didn't quite know what to say: "Well," he mumbled, "my father smokes, my grandfather smoked . . ."

The elder turned back to the young men: "See," he exulted, "the foreigner shows respect for his father and his forefathers. They smoked, so he smokes. Follow his example, show respect for your forefathers: they never smoked a cigarette; you shouldn't either!"

As this example illustrates, the same reasons that can be used to justify oneself can also be used as an argument to convince others. (The extra twist here is that when combined with other considerations, the same reasons can be used to justify one individual's practice and to argue against others adopting it: the import of reasons depends on the context.)

Retrospective and Prospective Uses of Reasons

Sometimes people use reasons to explain or to justify decisions already taken and beliefs already held. This is a retrospective use of reasons. Sometimes people use reasons as arguments in favor of new decisions or new beliefs. This is a prospective use of reasons. When reasons are used prospectively, it may

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

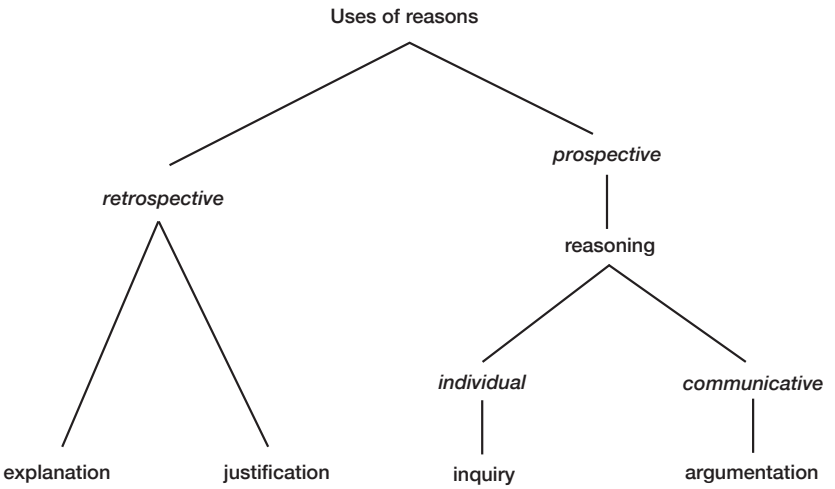

Figure 13. Retrospective and prospective uses of reasons.

be to answer a question of one's own to which one doesn't yet know the answer or to convince others of an opinion one already has. In the first case, it is inquisitive reasoning, and in the second case, it is argumentative reasoning. Figure 13 represents these uses of reason.

The philosophical and psychological literature treats retrospective and prospective uses of reasons as if they were altogether different topics, with little or no attention to what they have in common. We, on the contrary, will argue that there is no dividing line between retrospective and prospective uses of reasons, and that there are many cases of overlap. The use of reasons in explanation is often treated as more important than their role in justification (or, at least, as equally important). We, on the contrary, will argue that the justificatory role of reasons is more important than their explanatory role. The individual, inquisitive form of reasoning, aimed at answering questions on one's own, is considered reasoning par excellence, while the communicative, argumentative form is considered secondary. We, on the contrary, will argue that argumentation is primary.

Reasoning is a major topic in philosophy and psychology in its own right. The dominant view is that reasoning is aimed at truth and at good decisions and should be impartial and objective. The reasons used in reasoning should be impersonal "arguments" that owe their force to their formal properties

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

(studied in logic and probability theory). The use of reasons to explain and justify ideas already adopted or decisions already made and acted upon have, on the other hand, a personal component and cannot be really impartial.

How warranted is this contrast between impersonal reasoning and personal justification? Are personal considerations always irrelevant in good reasoning? In practical reasoning, at least, the relevance of personal considerations is generally recognized. Not surprisingly, philosophers working on practical reason and moral philosophy, in particular Joseph Raz,[1] have been more open to a unified approach to what we call retrospective and prospective uses of reasons.

There is, anyhow, a simple argument in favor of approaching in a unified framework retrospective and prospective reasons. The argument is that all reasons that can be used in justification can be used in reasoning and conversely. If you doubt this, imagine, for instance, a reason that would be good enough to retrospectively justify the opinion that no one should smoke. Then, surely, this reason could be used prospectively to convince oneself or others that no one should smoke. Conversely, imagine that a math teacher asks her pupils whether there is a greatest prime number and helps them reason and rediscover the proof that there is not. Then, surely, that very same proof, which caused the pupils in the first place to come to the conclusion that there is no greatest prime, becomes thereafter an impeccable justification for their belief, as it already was for the teacher.

In standard cases of argumentation, that is, in the production of reasons to convince others, the same reasons have both retrospective and prospective relevance. The arguer presents herself as trying to convince the addressee of an opinion she already holds. Olav and Livia, for example, are about to order raw oysters, and Olav wonders which wine they should have. He asks Livia, and she answers: "A Muscadet! It has just the right acidity and minerality to go with oysters." Olav has heard bad things about Muscadet, but given that Livia is more knowledgeable than he is about wine, her argument convinces him that they should have Muscadet and that he should revise his negative opinion. Livia's arguments to convince Olav are at the same time reasons that justify her own opinions. A sincere arguer uses as arguments to convince her audience reasons that she thinks provide good, retrospective justifications of her own views.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Not only do retrospective justification and prospective reasoning overlap in many ways, not only do they draw on the same pool of reasons; they also rely, we want to argue, on one and the same mechanism, a module that delivers intuitions about reasons.

Reasons Themselves Must Be Inferred

It would be quite surprising (and interesting) to find animals other than humans that think about reasons. Reasons occupy an important place in human thinking because, we have suggested, of the unique role they play in humans' very rich and complex social interactions. Reasons help establish personal accountability, mutual expectations, and norms. Saying this, however, doesn't tell us how humans are capable of knowing their reasons (even if this is a quite imperfect knowledge, as we have seen).

As we pointed out in Chapter 7, it takes more to have a reason than to just recognize some fact. You can walk out and see that the pavement is wet, but you cannot just *see* that this is a reason to believe that it has been raining. That the pavement is wet may be an objective reason for concluding that the pavement is slippery, that the outside temperature is not below freezing point, that one's shoes will get dirty, and so on. To think, "The pavement is wet" is not by itself to entertain a reason. You may, moreover, intuitively infer that it has been raining from the fact that the pavement is wet without this relationship between premise and conclusion being mentally represented in the process. Only if you were to entertain a thought like "From the fact that the pavement is wet it follows that it must have been raining" would you be recognizing the reason for your conclusion.

Suppose you do entertain a reason for inferring that it has been raining. The question still arises: *How* did you come to know that the fact that the pavement is wet is a reason for your conclusion? Reasons do not appear in our head by magic. Recognizing that some fact is a reason for inferring a given conclusion can only be achieved through—what else?—another, higher-order inference.

So, how are reasons inferred? By finding further reasons for our reasons? Sometimes, yes; most of the time, no. Assuming that the recognition of a reason must itself always be based on a higher-order reason would lead to an infinite

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

regress: to infer a conclusion *A,* you would need some reason *B;* to infer that *B* is a reason to infer *A,* you would need a reason *C;* to infer that *C* is a reason to infer that *B* is a reason to infer *A,* you would need a reason *D;* and so on, without end. Hence, the recognition of a reason must ultimately be grounded not in further reasoning but in intuitive inference. This infinite regress argument is an old one: in 1895 Lewis Carroll (of *Alice in Wonderland* fame) published an early version of it in a short and witty note entitled "What the Tortoise Said to Achilles."[2] But there is a question that had not been addressed: What implications, if any, does all this have for psychology?

Inferences, we have argued, are made possible by the existence of regularities in the world (general regularities like the laws of physics or local regularities like the bell–food association in the lab where Pavlov kept his dogs). A regularity that makes an inference possible need not be represented as a premise in the inferential process; it can instead be incorporated in a dedicated procedure. Intuitive inferences are produced by autonomous modules using such procedures. In this perspective, the fact that the recognition of reasons is grounded in intuitive inference suggests that there must be some regularity that a module can exploit in order to recognize reasons. Is there really such a module? If so, what is the regularity involved? Is there some better alternative account to explain how reasons are identified? In the psychology of reasoning, these issues are not even discussed.

Psychologists have studied the format in which people represent premises and the method by which they infer conclusions from these premises (this is what the debate between "mental modelers" and "mental logicians" that we evoked in Chapter 1 was all about). Much less studied is when and how people infer that specific premises provide a reason for a specific conclusion. The fact that reasons must ultimately be grounded in intuitive inference has been either ignored or deemed irrelevant, as if this ultimate intuitive grounding of reasons were too far removed from the actual processes of reasoning to be of consequence to psychology. Actually, just the opposite is the case.

In everyday reasoning, higher-order reasoning about reasons is quite rare. Most of the reasons people use are directly grounded in intuitive inference. It is intuitively obvious, for instance, that the pavement being wet is a reason to infer (with a risk of error) that it has been raining. When the intuitive grounding

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

of reasons is indirect, the chain is short. A fastidious reasoner might add one extra step: intuitively, the most likely explanation of the pavement being wet is that it has been raining; hence the pavement being wet is a reason to infer that it has been raining. Even in more formal reasoning, where people do reason about reasons, the intuitive ground is never very far. What makes reasoning possible, not just in principle but in practice, is the human capacity to intuitively recognize reasons.

The whole dual process approach of Evans, Kahneman, Stanovich, and others that we considered in Chapter 2 has at its core the assumption that intuitive inference and reasoning are achieved through two quite distinct types of mechanisms. We disagree. One of the main claims of this book is that *reasoning is not an alternative to intuitive inference; reasoning is a use of intuitive inferences about reasons.*

What makes humans capable of inferring their reasons is, we claim, their capacity for metarepresentational intuitive inference. To articulate this claim, we revisit and expand ideas introduced in Chapters 3 through 5 about three topics essential to understanding the human mind: intuitions, modules, and metarepresentations.

Intuitions about Reasons

Intuitions, we suggested in Chapter 3, are produced neither by a general faculty of intuition nor by distinct types of inferential process. They are, rather, the output of a great variety of inferential modules, the output of which is to some degree conscious while their operations remain unconscious. Our question now is: Is there a module that draws intuitive inferences about reasons? To answer, we must first sharpen our understanding of intuitions generally.

The inferential mechanisms that produce intuitions, that is, conscious conclusions arrived at through unconscious processes, are quite diverse. There is no intrinsic feature of intuitive inference that they would all share among themselves but not with other types of inference. By way of illustration, here are two cases of inference that are clearly both intuitive but that otherwise have very little in common.

Drawing on earlier work by Wolfgang Köhler (one of the founding fathers of Gestalt psychology), the neuroscientist Vilayanur Ramachandran and the

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

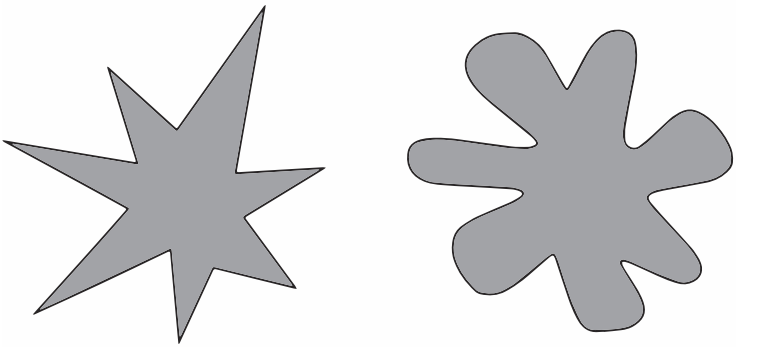

Figure 14. Kiki and Bouba.

psychologist Edward Hubbard presented people with two shapes from Figure 14.

They told them, "In Martian language, one of these two figures is a *bouba* and the other is a *kiki.* Try to guess which is which." Ninety-five percent of people picked the left figure as *kiki* and the right as *bouba.*[3] The strong intuition that it should be so is based, it seems, on a synesthetic association between sounds and shapes. Here intuition is close to perception.

The *bouba–kiki* intuition is quite concrete. Other intuitions are quite abstract. The English philosopher G. E. Moore noted in 1942 that it would be absurd to make a statement of the form "*P,* but I don't believe that *P*" such as, "It is Monday, but I don't believe that it is Monday." This observation is intuitively obvious, but it is not that easily explained. Contrary to what may seem, the proposition isn't self-contradictory: it could very well be true that it is Monday and that I don't believe that it is. The absurdity isn't in what is stated but in its being stated in the first person and in the present tense. That much is clear, but it isn't enough to explain the intuition of absurdity. In fact, while the intuition of "Moore's paradox" is uncontroversial, its explanation remains to this day a topic of controversy among philosophers.[4]

As the *bouba–kiki* and the Moore paradox examples illustrate, what renders some conclusions intuitive is neither their content nor the way in which they are produced. It is the kind of confidence we have in these conclusions. Intuitions are distinguished not by their cognitive features, but by their *meta*-cognitive features.[5]

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Confidence in our intuitions is unlike confidence in our perception or in our memory. Perception is experienced as a direct registration of how things are. Correctness of perception is generally taken for granted. Similarly, the way we use memory at every moment of our life is experienced as the direct recall of information that had been mentally stored. When perception and memory work fluently and unhampered, we are wholly unaware of the inferential work they involve.

We experience intuitions, on the other hand, as something our mind comes up with rather than as information that we just pick up from the environment or from memory. Our confidence in our intuitions is confidence in our mind's ability to go beyond the information given, in other terms, to draw inferences. It is not just that our intuitions feel right; we feel that *we* are right in coming up with them. The conclusions of intuitive inferences are experienced as personal thoughts. When we think of objective reasons to justify our intuitions, we readily assume that we must have had these objective reasons in mind to produce these intuitions. This sense of cognitive competence needn't be conceptually articulated. It may be just a metacognitive feeling that is activated only to the degree to which we pay attention to our intuition. Still, we claim, it is this distinctive kind of self-confidence that characterizes intuitions.

We have rejected the old idea that intuitions are the outputs of a distinct faculty of intuition and the currently fashionable idea that they are the outputs of a system 1 type of mechanisms of inference. This raises a puzzle. If, in and of themselves, intuitions are neither similar to one another nor distinct from other types of inference, why should we group them together and distinguish them from other mental states at all? Why do we have this special form of self-confidence that makes us set apart as intuitions some inferences that do not have that much in common otherwise?

Why indeed do we have intuitions at all? Here, as an aside, is a speculative answer. To distinguish a thought of your own as an intuition is to take a stance of personal authority on the content of that thought. This stance, we suggest, is less relevant to your own individual thinking than it is to the way in which you might communicate that thought to others. An intuition is a thought that, you feel, you may assert on your own authority, without an argument or an appeal to the authority of a third party. To make an assertion (or propose a

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

course of action) on the basis of your intuition is a social move that puts others in the situation of having either to accept it or to express distrust not just in what you are saying but in your authority for saying it. By expressing an intuition as such, you are raising the stakes: you stand to gain in authority if it is accepted, and to lose if it is not. Even if your assertion is rejected, however, putting it forward as an intuition of yours may help you withstand the authority or arguments of others. Intuition may license stubbornness, which sometimes is a sensible social strategy.[6]

Metacognition can take not only the simpler form of evaluative feelings but also the more elaborate form of metarepresentational judgments.[7] Say you are asked how long it will take you to finish writing the paper you have been working on. You reply, "A couple of days," but your feeling of self-confidence isn't very strong. Thinking about it, you are more confident that finishing the paper will take you *at least* a couple of days. What is involved now is more elaborate than a mere metacognitive feeling. You are metarepresenting two representations (*I will have finished writing the paper in a couple of days,* and . . . *in at least a couple of days*) and comparing your relative confidence in them. This time, metacognition is also metarepresentational. Distinguishing mere metacognitive feelings and more elaborate metacognitive metarepresentations is, we will show, essential to understanding reasons.

Our intuitions result from inferences about an indefinite variety of topics: *bouba* and *kiki,* Moore's paradox, the mood of a friend, or what film we might enjoy. Our metarepresentational abilities make us, moreover, capable of having intuitions about our intuitions.

Metarepresentational intuitions about our first-order intuitions may be focused on various aspects of these intuitions: they may be, for instance, about the reliability of first-order intuitions or about their acceptability to other people with whom we would like to share them. Some of our metarepresentational intuitions are not just about our degree of confidence in our first-order intuitions but—and this is crucial to the present argument—they are about the *reasons* for these intuitions.

You arrive at the party and are pleased to see that your friend Molly is there too. She seems, however, to be upset. When you have a chance to talk to her, you say, "You seem to be upset tonight." She replies, "I am not upset.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Why do you say that?" Just as you had intuited that she was upset, you now intuit reasons for your initial intuition. Here are what your two intuitions might be:

> *First-order intuition*: Molly is upset.
> *Metarepresentational intuition about your reasons for your first-order intuition:* the fact that Molly isn't smiling and that her voice is strained is what gives me reasons to believe that she is upset.

You want to go to the cinema and hesitate between *Superman 8* and *Star Wars 12*. You intuitively decide to go and see *Superman 8*. The film turns out to be rather disappointing, and you ask yourself why you had made that choice. An answer comes intuitively:

> *First-order intuitive decision*: To go to see *Superman 8*.
> *Metarepresentational intuition about your reasons for your first-order intuitive decision*: The fact that you had enjoyed *Superman 7* more than *Star Wars 11* was your reason for deciding to go to see *Superman 8* rather than *Star Wars 12*.

We typically care about our reasons when our intuitions are challenged by other people or by further experience.

We have intuitions not only about our reasons for our own intuitions but also about other people's reasons for their intuitions. You have spent a few hours inside, in a windowless conference room; you are now walking out of the building with your colleague Lin. The sky is blue, as it was when you walked in, and yet Lin says, "It has been raining." What reasons does he have to say that? Yes, it is a bit cooler than you might have expected, but is this enough to justify Lin's intuition? You look around and see a few puddles. Their presence, you intuit, provides a reason for Lin's assertion:

> *Lin's intuition:* It has been raining.
> *Metarepresentational intuition about Lin's reasons for his intuition:* The fact that there are puddles is Lin's reason to assume that it has been raining.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

If, moreover, you think that the reason you attribute to Lin is a good reason, you don't have to accept his assertion just on his authority; you now share a reason to accept it.

The Reason Module

What is this mechanism by means of which we intuitively infer reasons? What is the empirical regularity that makes reasons identifiable as such in the first place? While the attribution of reasons is hardly discussed, the attribution of beliefs and desires has been a central topic of debates in philosophy and psychology. As we saw, a common view—about which we expressed reservations—is that the regularity that helps us attribute beliefs and intentions to people is the presumption that they are rational beings. Couldn't such a presumption of rationality help us identify people's reasons? Should one attribute to others reasons that make their beliefs and intentions rational?

And what about our own reasons? How do we come to know them—or think we know them? Are the identification of other people's reasons and that of our own reasons based on one and the same mechanism, or on two distinct mechanisms operating in quite different ways?

As we saw in Chapter 7, there are good grounds to reject the idea that there is a power or a faculty of introspection that allows us to directly read our own mind. Moreover, it is not that we had reasons in mind when reaching our intuitive conclusions, reasons that we might then introspect if there were such a thing as introspection. Rather, we typically construct our reasons after having reached the conclusions they support. In order to attribute reasons to ourselves, then, we have to infer them, just as we have to infer the reason we attribute to others. Of course, we typically have much richer evidence about ourselves. We have some degree of direct access to our sensations and feelings. We can talk to ourselves in inner speech. Notwithstanding these differences in the evidence available, the way we draw inferences about our own reasons should be, in essential respects, similar to the way we draw inferences about the reasons of others.

Does this mean that, when we attribute reasons to ourselves or to others, we must be using *as a premise* the presumption that we (or they) are rational and tend to come to conclusions for which we (or they) have good reasons?

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

No, from a modularist point of view, regularities—such as the regularity that humans tend to think and act rationally—can be exploited by modular procedures without being represented as premises at all. We may, in other terms, exploit properties of human rationality in order to attribute reasons to others or to ourselves without having to entertain any general thought about human rationality itself.

There is both something appealing and something quite problematic about giving a central role to rationality in the discovery of reasons (whether in a standard or in a modularist perspective). What is appealing is the highlighting of a close link between reasons and rationality. It seems commonsensical that if we weren't rational, we wouldn't understand reasons, let alone care for them. If reasons weren't rational to some sufficient degree, we would not even recognize them as reasons.

What is much more problematic is the idea that a general human tendency to think rationally (together with some specific information about each case) should be enough to guide us in identifying reasons. For any rational belief or intention, there is an indefinite variety of quite different possible reasons, all compatible with the kind of limited evidence that might be available, which would rationally justify it.

Perhaps Lin's assertion that it had been raining wasn't based at all on the observation of the puddles. For all you know, it might have been based on his superior competence at recognizing and interpreting changes in the air's temperature and humidity; or perhaps he had noticed, while you were both in the conference room, someone entering with a wet umbrella; or he may have had quite different reasons that you have no clue about. What this means is that just looking for possible objective reasons that might have been accessible to a person doesn't come near being an effective heuristic to discover that person's actual reasons.

When we talked about mindreading in Chapter 6, we encountered a similar problem. Rationality doesn't by itself provide an adequate basis to attribute beliefs and intentions. The attribution of such mental states, we argued, takes advantage of the modular organization of the mind. What it exploits is not rationality as a general feature but the specific features of various cognitive competencies that humans share (and that jointly make them rational beings). This, we now suggest, is also true of the attribution of reasons. To attribute

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

reasons to others or to themselves, humans rely less on their overall rationality than on the effectiveness of some specific competencies.

Back to Molly at the party. She challenges you to explain why you said she was upset. To answer, you have to draw a backward inference from your initial intuition about her mood to reasons that would justify your intuition. You do not know for sure what might have triggered this intuition, but you are well equipped to make plausible assumptions.

You are not just a rational being; you have the typical human ability to recognize emotions. You have a well-developed expertise to use facial expressions, tone of voice, bodily movements, and so on as evidence of mood, allowing you to "read" the mood of others. While you cannot remember the cognitive process that led to your intuition—processes of intuitive inference are essentially opaque—you remember things you noticed when you met Molly this evening and that are standard indicators of moods: she was not smiling, her tone of voice was strained, and so on. So, you pick from what you remember (or from what you are presently noticing) pieces of evidence that might best justify your intuition that Molly is upset; you then infer that these were the reasons for your intuition.

Typically, your intuition about a person's mood is triggered by a combination of factors, many of which you are not even aware of, and that each makes some contribution to your intuition in the context of all the others. When you single out some pieces of evidence as being the reasons for your intuition you are typically exaggerating their weight as evidence, but this may be the condition for producing a relevant narrative.

Your memory, as we saw in Chapter 3, is not a mere recall of past registrations. It is constructive, and it often "remembers" features that help make better sense of what happened, even when, in fact, you hadn't observed these features at the time. It may be, for instance, that you hadn't noticed that Molly's tone of voice was strained until she asked you why you thought she was upset. Still, because it fits so well with your intuition that Molly was upset, this feature is injected into your memory of what you think caused your intuition in the first place. The strength of the reasons you now invoke is itself inferred. It is inferred from the confidence you have in your own intuition: if your intuition feels right, then your reasons for this intuition must be strong. You look for plausible strong reasons and assume that they are the reasons that motivated you.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

We intuitively infer our reasons for some specific intuition not on the general presumption of our own rationality, but on a much narrower confidence in the specific kind of competence that produced this intuition. Our feeling of rightness when we intuit the mood of a friend is based on our sense not that we are rational beings but that we are competent at judging people's mood and particularly the mood of people we know well.

When you have to infer the reasons that led *another* person to a given conclusion, your task is, again, to draw a backward inference from this conclusion to the kind of reasons that could explain and, at least to some extent, justify it.[8] When you walked out of the building with Lin and he said, "It has been raining," you looked around and, seeing puddles, you assumed that their presence provided a good reason for Lin's assertion. Why do you intuit that the puddles provide a reason to infer that it has been raining? Because, just like Lin, you have the competence to recognize evidence of the weather. When Lin's statement causes you to pay attention to the telltale puddles, you yourself intuit that it must have been raining. Since you trust your own inference from the puddles to the rain, you make the higher-level inference that the presence of such puddles is a good reason, for others or for yourself, to conclude that it just rained, and you further infer that this may well have been Lin's reason.

What about cases where we don't share others' intuitions? Suppose that when Lin said it had been raining, you saw the puddles but didn't intuit that it *must,* only that it *might* have been raining. In fact, paying more attention, you notice that the puddles—and there aren't that many—are all in the same limited area in front of you. You are now more disposed to infer that just the grounds in front of you happen to have been watered and that, no, it didn't rain. Still, puddles did evoke in you the possibility of rain: they were a prima facie reason to infer that it might have been raining, but it turns out not a good enough reason. Hence your intuition would still be that Lin's reason for stating that it had been raining probably was that he had noticed the puddles, but you would now judge this to be a poor reason. If, on the other hand, no possible evidence of rain had come to your mind, then you might have no intuition about Lin's reasons. You might just be and remain puzzled by his statement.

The attribution of reasons, to others or to oneself, needn't be more than a rather superficial affair. One searches the environment or memory for some actual or plausible piece of information (Molly's strained voice, puddles) that

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

could be invoked both to explain and to justify an intuition. If such a piece of information is found, it is assumed to be the actual reason for one's own intuition, a probable reason for someone else's.

Contrary to the commonsense view, what happens is not that we derive intuitive conclusions from reasons that we would somehow possess. What we do, rather, is derive reasons for our intuitions from these intuitions themselves by a further process of intuitive backward inference. We infer what our reasons must have been from the conclusions we intuitively arrived at. We typically construct our reasons as an after-the-fact justification.

We attribute reasons to others in the same way: to the extent that we trust their competence, we tend to trust their intuitions and to infer their reasons through the same process of backward inference. When we don't trust their competence, a similar process of backward inference will settle for apparent reasons that we ourselves find too weak or flawed to justify their intuitive conclusions but that they may have found good enough. When we believe others to be mistaken, we are typically content to attribute to them blatantly poor reasons.

We infer our reasons so that they should support our intuitive conclusions. We assess the strength of other people's reasons on the basis of our degree of agreement with their conclusions. Does this mean that this search for reasons is a purely cosmetic affair, a way of dressing up our naked biases just to look good in our own eyes and to have others look good or bad depending on whether we agree or disagree with them? Is there no cognitive benefit to be expected from the process? No, this wouldn't make much sense. If everybody just stood by their initial intuitions, come what may, reasons would be altogether irrelevant.

Reasons, we have argued, are for social consumption. People think of reasons to explain and justify themselves. In so doing, they accept responsibility for their opinions and actions as justified by them; they implicitly commit themselves to norms that determine what is reasonable and that they expect others to observe. In giving reasons, people take the risk of seeing their reasons challenged. They also claim the right to challenge the reasons of others. Someone's reputation is, to a large extent, the ongoing effect of a conversation spread out in time and social space about that person's reasons. In giving our reasons, we try to take part in the conversation about us and to

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

defend our reputation. We influence the reputation of others by the way we evaluate and discuss their reasons.[9]

So, no, we don't invoke reasons for some inane ego-boost, but, yes, the very way we infer our reasons is biased in our favor. We want our reasons to justify us in the eyes of others. Because they are going to be submitted to others' judgment, reasons may be rethought and revised to be better accepted. Sometimes this means revising, moreover, the conclusions that our reasons support: changing opinion or course of action so as to better be able to justify ourselves. Reasons and conclusions may, in the end, have to be mutually readjusted.

There is, we are assuming, a dedicated metarepresentational module, the job of which is to infer reasons, ours and those of others. Its job is not to provide a psychologically accurate account of the reasons that motivate people. In fact, the implicit psychology—the presumption that people's beliefs and actions are motivated by reasons—is empirically wrong. Giving reasons to justify oneself and reacting to the reasons given by others are, first and foremost, a way to establish reputations and coordinate expectations.

Does it follow that the reasons we give and expect others to give are merely adjusted to some local consensus, some culturally constructed notion of rationality, and that one shouldn't expect people's reasons to be rational in an objective sense of the term? No, this doesn't follow at all. The reasons people give play a much more important role than just signaling that they are norm-abiding members of their social group. People get the good reputation they care about when they are seen as reliable sources of information and as effective partners in cooperation. There is no way they could maintain over time such a reputation without the basic kind of objective rationality that makes them draw cognitively sound inferences and act effectively. To serve their reputational purpose, the personal reasons people invoke should be recognized by others as representing objective reasons, and the best way to secure this recognition is, at least, to invoke reasons that, objectively, are good or at least not too bad.

A cultural community may favor certain types of reasons such as reliance on specific authorities. It may unequally recognize the competence of women and men, young and old, socially inferior and superior, in invoking reasons. It may condone some irrational reasons, such as premonitory dreams. What a community cannot do is build a battery of reasons all of its own. Everywhere,

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

people's intuitions about reasons are anchored in cognitive competencies that, to a large extent, they share as members of the human species, competencies that contribute to humans' cognitive efficiency, that is, rationality in a basic sense of the term. Without such cognitive anchoring, we doubt that any norms of rationality could ever emerge and be maintained in a social group.

Our reasons tend to be rational because, in the first place, our intuitions tend to be rational. What humans do and, presumably, other animals don't do is add to their spontaneous inferences higher-level representations of reasons for these inferences. These reasons are not what makes human inferences rational in the biologically relevant sense of cognitively efficient. What these reasons help do, rather, is *represent our inferences as rational* in a different, socially relevant sense of the term where being rational means, precisely, being based on personal reasons that can be articulated and assessed.[10] The public representations of beliefs and intentions as guided by personal reasons are a fundamental aspect of human social interaction. These representations, we suggest, are produced by a dedicated metarepresentational module. All our reasons are, directly or indirectly, outputs of this module.

Could, then, human Reason (with the capital *R* used in classical philosophy) be a module? Have we found it? Should the next step be the localization of the Reason module in the brain? No, no, and no. Classical ideas about Reason are not about a psychological mechanism but about an essential and transcendental feature of the human mind as a whole. In any case, we have not *found* any module; we are merely speculating, with, we hope, sensible arguments, that the identification of reasons might well be the job of a dedicated module. If we are right and there is such a module, it would have indeed to be realized in some neural structure, but that structure needn't occupy—and occupy alone—a single locus in the brain. In any case, such a module for inferring reasons wouldn't correspond to Reason as classically understood.

Descartes and many other philosophers have sung praises of Reason (while other thinkers have been less lyrical). The module we are talking about, if it exists, would not be something humans would want to brag about as they have bragged about Reason. Still, the closest thing to classical Reason to be found in the human mind/brain may well be this module. We will therefore call it the *reason module* with a modest, lowercase *r*.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Can the Reason Module Reason?

The reason module is, at least in part, aimed at producing justifications and is very much biased in our favor. How could the reasons it produces ever improve on our intuitive inferences if they are inferred from them through backward inference? How could it help in reasoning? How could our evaluation of other people's reasons ever be more than a projection of our self-serving prejudices? How could we ever be convinced by the reasons of others to change our own views?

Part of the answer is that our first-order intuitions (about Molly's mood, the rain, and the vast variety of things about which we have such intuitions) are delivered by a great many modules, while our metarepresentational intuitions about reasons for our first-order intuitions are delivered by one metarepresentational module that just works on reasons. First-order modules draw all kinds of inferences about objects in their domain of competency by exploiting regularities in their domain. The metarepresentational module involved draws inferences in its domain of competence; it draws, that is, inferences about the relationship between reasons and conclusions. To do so, it attends to relevant properties of this relationship. Some intuitions may come easily to us and feel quite strong, but finding reasons that feel as strong as these intuitions themselves may not come so easily. Low confidence in reasons for an intuition may undermine initially high confidence in that intuition.

We may, for instance, have an evolved disposition to accept important risks for exceptional opportunities. Such a disposition, possibly advantageous on average in a distant ancestral past, may, in the modern world, easily be exploited by swindlers.

Jeb, for instance, has an immediate intuition that if he responds favorably to a message he just received from the widow of a rich banker asking him to help her transfer millions of dollars to his country, he will become extremely rich, but then he may have trouble finding credible reasons—reasons that he could share with his family and friends—in support of this intuition. He may initially dismiss his friend Nina's warning that this is a scam, but then he might soon enough find good reasons for her warning. His skeptical intuitions about reasons are different from and better than his enthusiastic intuition about his good fortune.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Others often express intuitions that differ from ours. When trying to explain these intuitions that we do not share, we may nevertheless intuitively attribute to them reasons that we find good or even compelling, leading us to revise our initial intuition. A friend of Nikos and Sofia asks them to solve the following problem (actually, a problem much studied in recent psychology of reasoning).[11] A bat and a ball together cost 1.10 euros. The bat costs one euro more than the ball. How much does the ball cost? For Nikos, the intuitive answer is that the bat costs ten cents. To his surprise, Sofia answers, "The ball costs five cents." Seeing him puzzled, she continues, "If the ball costs five cents, then the bat must cost 1.05 euros . . ." Before she finishes, he sees it: the difference is one euro, as required! This, he intuits, must be Sofia's reason for her answer, and it is a good reason. He therefore rejects his own initial intuition that the ball must cost ten cents and accepts Sofia's answer.

The strength of our first-order intuitions and that of our corresponding metarepresentational intuitions need not match. Reasons may strengthen or undermine our first-order intuitions, and sometimes lead to revisions. Reasons, then, need not be mere stamps of approval on our first-order intuitions.

Don't, however, take Jeb's or Nikos's example as typical. Jeb's initial intuition went against common wisdom. It was clear enough that he would have to justify himself if he acted on it; plausible justifications were hard to come by. And then he had Nina's help in finding reasons to reconsider. Similarly, Nikos might not have revised his solution to the bat-and-ball problem if Sofia had not come up with a different solution and started explaining it. Few of our intuitions are as blatantly stupid as Jeb's or as demonstrably false as Nikos's. Most of our first-order intuitions are at least plausible, and backward inference usually yields plausible second-order reasons to justify them.

Unless our reputation is at stake, we are unlikely to seriously examine our own first-order intuitions in the light of our metarepresentational intuitions about reasons. Even if we do, the reasons that come easily to our mind are likely to confirm or even strengthen our initial intuitions. And even if there is a mismatch between our first-order intuitions and our second-order intuitions about reasons, we may not automatically trust the latter more than the former. Victims of scams like Jeb may have some higher-order doubts but fall into the trap all the same: first-order intuitions are too strong. So the fact that our

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

first- and second-order intuitions might not match doesn't, by itself, make us wise.

We began this book with a double enigma, the second part of which was: How come humans are not better at reasoning, not able to come, through reasoning, to nearly universal agreement among themselves? It looks like now we might have overexplained why different people's reasons should fail to converge on the same conclusion and ended up with the opposite problem: If the reason module is geared to the retrospective use of reasons for justification, how can it be used prospectively to reason? How come humans are capable of reasoning at all, and, at times, quite well?

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

9

Reasoning: Intuition and Reflection

Reasons can justify an opinion already formed or a decision already made—this is the retrospective use of reasons. But what if you have a question in mind that you don't know how to answer, a decision to make but you are not sure which? Isn't this where a prospective use of reasons—reasoning proper—should help? In principle, yes. If your intuitions on the issue are not clear or strong enough to sway you one way or another, nothing could be better, it seems, than to think in an impartial way of reasons showing which answer is right or which decision is best. Is this, however, what the reason module helps you do?

Intuitive Arguments, Reflective Conclusions

The intuitions the reason module provides are not, we have stressed, about facts that could be a reason for some unspecified conclusion; their form is not "*P* is a reason" (for example, "That Amy has a fever is a reason"). These intuitions are about facts taken together with the conclusion that they support; their form is "*P* is a reason for *Q*" (for example, "That Amy has a fever is a reason to call the doctor").

In both justification and reasoning, the output of the reason module is a higher-order conclusion that there are reasons for a lower-order conclusion. In the case of justification, the lower-order conclusion has already been produced and possibly been acted upon; say, Amy's parents have already called the doctor. What is now added is the higher-order conclusion that Amy having a fever justified this decision. This retrospective justification may work even

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

if Amy's parents called the doctor just because she was feeling dizzy, before they actually took her temperature.

In reasoning, the reason module produces not one but two new conclusions, the second conclusion embedded in the first. The first conclusion is the higher-order argument itself, that is, the metarepresentational intuition that such-and-such reasons support a particular conclusion: Amy's fever is a good reason to call the doctor now. The second new conclusion is the conclusion—Let's call the doctor!—embedded in the overall argument and supported by it.

If you catch a fish that has just swallowed another fish whole, you catch two fish at once. Similarly, if you infer that some new conclusion is supported by good reasons, you infer two new conclusions at once: the whole argument and the conclusion the argument supports. When the embedded conclusion is relevant on its own, as is generally the case in reasoning, you may disembed it and assert it, store it in memory, or use it as a premise in further reasoning.

Two illustrations:

You are quite confident that you have left a book you need in your study, but however hard you look, you fail to find it. You try to think where else it might be, and you think, yes, it might also be in the living room. At this point, you reach a conclusion in the form of an argument:

> *Argument:* Since I'm pretty sure the book is either in the study or in the living room and I cannot find it in the study, I should look for it in the living room.

This is the kind of argument one typically accepts without considering higher reasons for it. Given that an intuition, simply defined, is a conclusion accepted without attention to, or even awareness of, reasons that support it, your argument as a whole is definitely an intuitive conclusion, an intuition. This intuitive conclusion, however, is *about* reasons and about the support these reasons give to a second conclusion, which is embedded in the argument:

> *Embedded conclusion:* I should look for the book in the living room.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Is the conclusion that you should look for the book in the living room also an intuition? No, because it doesn't fit the definition of an intuition. It is, after all, supported by reasons, reasons that you very much have in mind. In fact, it occurs embedded in a representation of these very reasons.

If the conclusion embedded in an argument is not an intuitive conclusion, what kind of conclusion is it? It is a reasoned conclusion, or, to use a term common in the literature on reasoning and that we have often used ourselves,[1] a *reflective conclusion,* a conclusion accepted because of higher-order thinking (or "reflection") about it. As this example illustrates, the amount of reflection that goes into reaching a reflective conclusion may be minimal. Still, there is a difference between intuitively believing, without any reasoning involved, that you should look for the book in the living room and having intuitive reasons for this conclusion.[2]

Here is a second example of a reflective conclusion, this time involving a little more reflection. You are sitting with a friend in a café. She challenges you: "I offer you two bets. Accept one of the two bets, and if you win it, I'll pay for the drinks." The two bets she offers are the following:

> *Bet A:* You win if at least three of the next five people to enter the café are males.
> *Bet B:* You win if at least three of the next six people to enter the café are males.

Without having to consider higher-order reasons, you conceive of an intuitive argument that helps you choose one bet.

> *Argument:* The chances of winning bet B are greater than the chances of winning bet A, making bet B the better choice.
> *Embedded conclusion:* Bet B is the better choice.

Your conclusion that bet B is preferable is supported by your intuition about your relative chances of winning the two bets. This intuition provides a reason for what is, therefore, a reflective conclusion.

What these two examples show is that a reflective conclusion need not be the output of a *mechanism* of reflective inference that could be contrasted to

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Table 1 Outputs of the reason module

	Types of conclusion	Form of conclusion
Direct output	Intuitive argument	Reasons $R_1, R_2, \ldots, R_n$ support conclusion C
Indirect output	Reflective conclusion	C

mechanisms of intuitive inference. A reflective conclusion may be an indirect output of a process of intuitive inference. The *direct* output of this process is an intuitive conclusion about reasons $R_1, R_2, \ldots, R_n$ for some conclusion C—in other terms, an argument; the *indirect* output is the reflective conclusion C. This is represented in Table 1.

We want to make an even stronger claim: not just some but *all* reflective conclusions in human thinking are indirect outputs of a mechanism of intuitive inference about reasons.

In the basic and most common cases (for instance, in the misplaced book or the choice of bet illustrations we just gave), a reflective conclusion is embedded in an intuitive argument. In slightly more complex cases, there may be not one but two levels of embedding, with an intuitive argument supporting a reflective argument that itself supports a final reflective conclusion.

Take the bet example. We described you as having the intuition that the chances of winning bet *B* are greater than the chances of winning bet A. This intuition would give you a strong argument for choosing bet B. You might, however, be more reflective and accept this argument in favor of bet B only on the basis of a higher-level intuitive argument:

> *Intuitive argument:* If only two of the next five people to enter the café are males, you would lose bet A, but you might still win bet B. All it would take is that the sixth person to enter the café be male. Hence, the chances of winning bet B are greater than the chances of winning bet A, making bet B the better choice.
> *Reflective argument:* The chances of winning bet B are greater than the chances of winning bet A, making bet B the better choice.
> *Reflective conclusion:* Bet B is the better choice.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

In this example, the higher-level argument (in which the reflective argument and the final conclusion are embedded) is, for most people, likely to be intuitive. For someone who doesn't find this more general argument intuitive or is unwilling to accept it without reflection, it could itself be embedded in an even more general argument about probabilities that would support it. There may be more than two levels of embedding. Still, ultimately, any reflective conclusion is the indirect output of a process of intuitive inference.

It would be a mistake to assume that higher-order explicit arguments make conclusions easier to understand and accept. To take an extreme example, for most people, it is a plain, incontrovertible fact that $1+1=2$. Alfred North Whitehead and Bertrand Russell famously devoted hundreds of pages of their landmark book, *Principia Mathematica*,[3] to deriving this conclusion through a long chain of complex arguments. Very few people are able to follow these arguments. They were not aimed, anyhow, at strengthening the rock-solid common intuition that $1+1=2$ but at demonstrating that it is provable and, in so doing, at providing logical foundations for mathematics.

What proportion of the many conclusions humans come to in their life are reflective rather than intuitive? What proportion of these reflective conclusions are themselves embedded in reflective rather than intuitive arguments? How common are higher and higher-order levels of reflection? These are empirical questions that have not been properly studied.

Note that the people who would be most likely to express with the confidence of experts their opinion on the importance of reflection in human inference are logicians, philosophers, and psychologists of reasoning. These experts are hardly typical of the population at large; they commonly resort to higher-level arguments as part of their trade. They might be prone, therefore, to mistake their own professional twist of mind for a basic human trait. We surmise that most human reasoning, even excellent reasoning of the kind that is supposed to make us humans so superior, rarely involves more than one or two levels of arguments.

We had suggested in Chapter 8 that the reason module produces, as direct output, intuitions about reasons and, as indirect output, reflective conclusions supported by these reasons. If this is right, then there is no need—and, in fact, no room—for a psychological mechanism ("system 2" or

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

otherwise), the job of which would be to directly produce reflective conclusions. Reasoning can be more or less reflective depending on the degree to which the arguments involved are themselves embedded in higher-order arguments. More reflective, however, does not mean less intuitive. From elementary reflective conclusions directly embedded in an intuitive argument to the many-level reflections characteristic of some of the most impressive scientific achievements, reasoning is always an output of a mechanism of intuitive inference.

In describing reasoning as a use of intuitions about reasons, we adopt a purely psychological approach quite different from more traditional logicist approaches to reasoning. Doesn't logic play a central role in reasoning? Aren't we missing something essential?

To better understand the issue, compare the psychology of reasoning to that of mathematics. The science of mathematics itself is about objective mathematical facts. The psychology of mathematics is about the way in which people learn about these facts and use them to calculate, for instance, accrued interests or the surface of a garden. The psychology of mathematics cannot be approached in a purely psychological way; mathematical facts have to be part of the picture. Similarly, it could be argued, reasoning is the use of objective logical facts. Intuitions about reasons that are not about their logical properties are no more part of reasoning than beliefs about, say, lucky numbers are part of mathematical thinking. If this argument is correct, then the psychology of reasoning should have logic at its core.

The acquisition and use of mathematical competence rely on the existence of mathematical symbols for numbers, operations, and so on. Even before the invention of writing, many words in spoken languages served as mathematical symbols and allowed a modicum of mathematical competence to develop. Logic, likewise, relies on symbols to represent propositions, logical relationships, and so on. The use of a whole range of specialized symbols for logic is relatively recent, but just as in the case of mathematics, many words and expressions of ordinary language can serve as logical symbols. Language, so the argument goes, made reasoning possible well before the development of more formal logic.

So, according to the classical approach to reasoning, language is essential to reasoning. We agree, but for completely different reasons.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Reasoning Relies on Language

Reasons, we have argued, are for social consumption. To be socially shared, reasons have to be verbally expressed and, indeed, reasons appear on the mental or public scene in verbal form. Reasons may serve to justify oneself, to evaluate others, or to convince people who think differently. All this involves verbal communication.[4] Even when you think of reasons on your own, often it is as if you were mentally answering what others have said or might say, as if you were readying yourself to question others' opinions and actions. Even when you think of reasons to answer your own questions, it is as if you were engaging in a dialogue with yourself. For this, you resort to inner speech.

Language is uniquely well adapted to represent reasons. To understand a reason is to mentally represent the relationship between at least two representations: the reason itself and the conclusion it supports; in other words, it is a metarepresentational task. Language and metarepresentation are closely associated (even if metarepresentation, and in particular mindreading, may well be possible without language).[5] Language is a uniquely efficient tool for articulating complex metarepresentations and for communicating them. Language, in particular, may be uniquely well suited to metarepresent relationships between reasons and conclusions.

Linguistic expressions can be embedded within linguistic expressions, and in particular, sentences can be embedded within sentences: "It is nice" is a sentence, and so is "Yasmina said that it is nice." Several sentences can be embedded in a more complex sentence to articulate a reason. For instance,

Molly isn't smiling.
Molly is upset.

are sentences that may represent states of affairs. They can be combined, as in

The fact that Molly isn't smiling is a reason to believe that she is upset.

This complex sentence metarepresents the relationship between a reason and the conclusion it supports.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

To express all kinds of metarepresentational relationships and in particular reason-conclusions relationships, language offers a variety of linguistic devices. Nouns such as "argument," "reason," "objection," and "conclusion" can describe statements and their relationships, and so can verbs such as "argue," "object," and "conclude." A number of so-called discourse markers, such as the connectives "so," "therefore," "although," "but," "even," and "however," have an argumentative function: they focus attention on a reason-conclusion relationship and give some indication of its character.[6]

Compare, for instance, these two possible answers to the question, "How was the party?"

a. It was a nice party. Pablo had brought his ukulele.
b. It was a nice party but Pablo had brought his ukulele.

Both answers state the same facts, but they put them in a different perspective. In the (a) answer, the information that Pablo had brought his ukulele would be understood as an elaboration and a confirmation of the statement that it was a nice party; the music contributed to the success of the party. In the (b) answer, on the contrary, what "but" does is suggest that some of the consequences you might have inferred from the assertion that it was a nice party, such as that everything went well, do not actually follow in this case: yes, it was a nice party, but this was in spite of rather than because of Pablo's music.

Argumentative devices such as "but" play a heuristic role in argumentation. What they do is facilitate inference and suggest which implications should and shouldn't be derived in the context. According to the classical view, however, the main role in verbal reasoning belongs to verbally expressed logical symbols—in particular, logical terms such as "or," "if . . . then," "only if," "and," and "not" (and other logical devices such as quantifiers and modals). Such logical devices are what make it possible to construct a valid deductive argument. They are the linguistic tools that make reasoning (or at least so-called deductive reasoning) possible. Really? Are the logical devices the protagonists of reasoning, and other linguistic devices mere supporting characters? We want to tell a different story.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

The Adventure of the Stolen Diamond and the Missing Premise

Contrary to received wisdom, sound syllogisms (and other sound formal deductions) do not provide compelling arguments in ordinary reasoning. We illustrate why not with a new Sherlock Holmes adventure.

A telegram is brought to Sherlock Holmes. He reads it, turns, and addresses his friend Dr. Watson:

> *Holmes:* The butler didn't steal the diamond. So, it is the gardener who did.

Watson, trusting Holmes's powers of deduction, might be convinced.

According to the standard logicist view, however, Holmes failed to express a proper deductive argument. Note that he used an argumentative device, "so," but no logical terms. At best, his statement is an enthymeme, that is, a truncated version of a true logical argument. If so, Holmes could have been more explicit:

> *Holmes:* Either the butler stole the diamond or the gardener did *[first premise].* The butler didn't *[second premise].* So the gardener stole the diamond *[conclusion].*

This logical version with the logical terms "or" and "not" corresponds to a disjunctive syllogism:

Premises:	1. *P* or *Q*
	2. not *P*
Conclusion:	*Q*

Such a syllogism is said to be *valid.* When the premises of a valid syllogism are true, the syllogism is said to be not only valid but *sound:* the conclusion of a sound syllogism is necessarily true. Perhaps Watson recognized the full syllogism that Holmes had expressed in a truncated form and was convinced not just by his trust in Holmes's logical acumen but by the actual logic of the argument.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Does, however, a syllogism that you know to be sound provide you, by itself, with a sufficient argument in favor of its conclusion? It is a common mistake to think so.

Suppose the telegram Holmes has just received is from the chief of police and reads: "The diamond was stolen by the gardener." Holmes, who until then had no idea who might have stolen the diamond, rightly accepts this as a fact, but rather than just report it to Watson, he decides to convey the information in syllogistic form. He might say the following, speaking truly:

> *Holmes:* Either the pope stole the diamond or the gardener did. The pope didn't. So the gardener stole the diamond.

Holmes's pseudo-argument is blatantly circular, and Watson should easily recognize this: since the pope was never considered a possible culprit, Holmes's only plausible ground for asserting the first premise of the argument ("Either the pope stole the diamond or the gardener did") is that he already knew that the gardener stole the diamond and could have said so directly. Sure, the syllogism is sound, but Watson would recognize it not as a genuine argument but as a somewhat quirky way for Holmes to assert what he knew independently. Watson's reason to accept that the gardener stole the diamond would, in that case, have nothing to do with the logical soundness of Holmes's pseudoargument and everything to do with his trust in Holmes.

In the same circumstances, a disingenuous Holmes could have been just as logical and strictly truthful but nevertheless misleading. He could have spoken as follows:

> *Holmes:* Either the butler stole the diamond or the gardener did. The butler didn't. So the gardener stole the diamond.

Replacing the pope with the butler doesn't change a bit the logic of the syllogism, nor does it alter its soundness, but it is likely that this time Watson would mistake it for a genuine argument, a reason to accept its conclusion. Why? Because, unlike the pope, the butler might have been suspected,

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

and Holmes might have deduced that the gardener stole the diamond from information that the butler didn't. Watson could think that Holmes had produced not only a sound syllogism but also a genuine argument.

More generally, for any fact you happen to know, you can always construct a sound syllogism that has this fact as its conclusion and that suggests that you came to know this fact thanks to your powers of deduction. Such a syllogism, however, doesn't give any genuine support to its conclusion. For a syllogism to serve as a bona fide argument in support of its conclusion, there must be good, noncircular reasons to accept its premises. This is not a logical requirement. From a logical point of view, circularity isn't a problem. Noncircularity, on the other hand, is a requirement of good reasoning. Hence, when a full syllogism is presented on its own as an argument, the argument itself is incomplete; it carries an implicit premise. It implies that there are good, noncircular reasons to accept the explicit premises and, as a consequence, the conclusion.

The common wisdom is that most arguments ordinarily used in conversation are, in fact, truncated syllogism (which they sometimes are, but much less often than is generally assumed). What we have just shown is that arguments consisting just of a syllogism, even a fully explicit one, are themselves truncated arguments: they implicitly convey that their premises are not just true but provide genuine reasons to accept the conclusion. There goes the alleged systematic superiority of explicitly laid-out syllogisms over informal statements of reasons: both, in fact, provide incomplete arguments.

In Reasoning, Logic Is a Heuristic Tool

Syllogisms are not better arguments for inquisitive or argumentative reasoning. They are an altogether different kind of thing.[7] Syllogisms (and deductions generally) are abstract formal or semiformal structures that make explicit a relationship of logical consequence between premises and conclusion. They may be used for a variety of purposes, but in themselves, they don't *do* anything. The arguments used in reasoning, on the other hand, are not defined by their structure, which is quite variable, but by what they do, namely, provide reasoners with reasons to come to some conclusion.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

It is a commonplace that merely recognizing a syllogism as valid is not a reason to accept its conclusion. We have shown, moreover, why recognizing a syllogism as sound isn't a sufficient reason to accept its conclusion either. Only if you have independent reasons to accept its premises may a sound syllogism give you a reason to accept its conclusion. What the syllogism does, in such a case, is help you see how the reasons in favor of the premises are also reasons in favor of the conclusion.

What if you take a given syllogism to be sound and you do have good noncircular reasons to accept the premises? Shouldn't you, in this case, accept the conclusion? The answer is again no. What you have now is a reason either to accept the conclusion or to change your mind about at least one of the premises. This is far from being a mere theoretical possibility or a rarity. Actually, a common use of syllogisms in reasoning is to bring people to revise their beliefs by showing them that these beliefs entail consequences that they have strong independent reasons to reject.

A reductio ad absurdum, as this form of argumentation is called, consists in arguing against a claim by showing that it leads to an absurdity, or at least to a manifest falsehood.[8] Suppose that Watson answered Holmes's apparent demonstration that the gardener was the thief by informing him, "My dear Holmes, your syllogism is, as usual, impeccable but your conclusion cannot be true: the gardener had a heart attack and died the day before the diamond was stolen; I happen to be the doctor who signed the death certificate!" Watson's testimony would override the words of the chief of police. The very logicality of the syllogism would then force Holmes to reject at least one of its two premises: perhaps the thief was neither the butler nor the gardener but someone else, or else the thief must have been the butler after all.

Most ordinary inquiries and arguments are about empirical facts. They involve premises that, even when we accept them as true, are less than certain and admit of exceptions. A conclusion derived from such premises inherits their precariousness. A syllogism with premises that are not strictly true is not a strict proof. As a result, having a good or even a compelling reason to reject the conclusion of a syllogism should cause one to reconsider the premises—to reconsider them, yes, but not necessarily to reject them completely. What may happen rather is that a premise, rather than being simply rejected, is recognized to have exceptions.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

To illustrate, we adapt the case of Mary having an essay to write (which Ruth Byrne used to argue that *modus ponens* deductions can be "suppressed," as we saw in Chapter 1).

Tang and Julia are Mary's flat mates. They wonder whether she will be back in time for dinner:

Tang: Well, if Mary has an essay to write, she stays late in the library.
Julia: Actually, she does have an essay to write. Let's not wait for her!

So they prepare the dinner, set the table, and open the wine, and then, just as they are about to sit down, Mary arrives.

Julia: We thought you had an essay to write and you would stay late in the library.
Mary: I do have an essay to write, and if the library had remained open, I would have stayed there.
Tang: A glass of wine?

Tang's and Julia's initial statements provided the two premises of a conditional syllogism, the conclusion of which was that Mary would stay late in the library (and hence not be back in time for dinner). They both initially accepted this conclusion but then, obviously, had to reject it when Mary appeared. Mary's explanation, however, showed that their premises had been credible, and their reasoning sensible; the problem was just that the circumstances were not quite normal.

A statement like Tang's, "If Mary has an essay to write, she stays late in the library," was not intended or understood to express a necessary truth or an undisputable empirical fact. Like most such statements in ordinary life, it expressed high probability in normal conditions—but then, conditions are not always normal. Tang had no need to qualify his statement by saying something like "probably" or "in normal conditions"; this is how Julia would have understood it anyhow. Mary's unexpected return shows that the two premises of the syllogism were not a sufficient condition to accept its conclusion—not because one of them was false, but because one of them admitted of exceptions, as do most of our ordinary life generalizations.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

So, it turns out, reasons to reject the conclusion of a syllogism do not even force you to reject any of its premise; they might just make you more aware of the in-normal-conditions (also called "ceteris paribus") character of at least one of the premises.

How to account for all this? One might deny that "if" and other connectives such as "and" and "or" have the logical sense classically attributed to them. Perhaps their sense is based on a nonmonotonic logic; perhaps it is probabilistic. These are semantic solutions: they consist in revising our understanding of the sense of words. There is, however, another way to go. The philosopher Paul Grice argued for an alternative, pragmatic solution that would explain what is happening with logical connectives (and, actually, with words and utterances in general) by focusing less on what words mean and more on what speakers mean when they use these words. He thought that the classical semantics of logical connectives could be preserved and that apparent counterexamples could be explained in pragmatic terms. (Of course, the semantic and the pragmatic approaches are not incompatible; on the contrary, in a full account, they should be integrated).[9]

Here we make a brief remark on the pragmatics of connectives and of syllogisms generally. Grice's insights have been developed in modern pragmatics and in particular in relevance theory. A basic idea of relevance theory is that the linguistic sense of words and sentences is used not to *encode* what the speaker means but merely to *indicate* it—indicate it in a precise way but with room for interpretation.

A tourist asks a Parisian a question:

Tourist: How far is the Eiffel Tower?
Parisian: It is near.

In the Parisian's answer, the word "near" may convey "within a short walking distance" if the tourist who is asking is on foot, and "within short driving distance" if the tourist is in a car. The word "near," of course, has a linguistic sense, but this sense is not identical to the meaning the speaker intends to convey; rather, this sense makes it possible to infer what the speaker means, given the context.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

A word such as "near" has a vague sense that must be made more precise in context. But what about words, such as "straight," that have a precise sense? These too serve as a starting point for inferring the speaker's meaning:

Tourist (standing in front of the American Church on the Quai d'Orsay): What is the way to the Eiffel Tower?
Parisian: Keep walking straight ahead. You can't miss it.

Actually the Quai d'Orsay is not a straight street. It follows the curves of the river Seine. Were the tourist to follow the advice literally and to keep walking straight ahead, she would end up in the river. In the context, however, "straight" conveys that she should stay on the same street, even if this makes for a curved rather than a straight trajectory. Here, the word "straight" is used loosely. So words can be used to indicate a meaning that is narrower or looser than the sense linguistically encoded.[10]

Logical connectives (as well as quantifiers and modals) behave like ordinary words. They don't encode the communicator's intended meaning but merely indicate it. Take the case of "or." The sense of "or" is such that a statement of the form "*P* or *Q*" is true if one of the two disjuncts *(P, Q)* is true. So, for instance, if the gardener stole the diamond, then Holmes speaks truly when he says, "Either the butler stole the diamond or the gardener did." But, as we pointed out, Holmes would be typically understood to mean not only what his utterance literally and explicitly means but also implicitly that he has some reasons to assert the disjunction other than just knowing that one of the disjuncts (that the gardener stole the diamond, in this case) is true. Typically, a "*P* or *Q*" statement conveys a greater confidence in the disjunction itself than in each of the disjuncts. Thus, in most ordinary contexts, the word "or" conveys more than the logical sense it encodes.

"Or" can also be used to convey less than its logical sense. Imagine the following dialogue:

Police chief: Either the butler stole the diamond or the gardener did.
Holmes: Are we sure? Either the butler stole the diamond or the gardener did, or some other member of the household did, or another inhabitant of the village, or of the county, or of the country.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

If it turned out that the thief wasn't even an inhabitant of the country but just a visitor from abroad, it wouldn't make much sense to say that Holmes had been wrong. He is clearly using a multiple disjunction (". . . or . . . or . . . or . . .") to express a kind of implicit meaning that utterances with "or" typically convey: that each of the individual propositions connected by "or" is in doubt, and he is doing so without committing to the truth of the whole disjunction itself. He is, in other terms, expressing both more and less than the literal sense of his utterance.

Other logical connectives such as "and" and "if" (as well as quantifiers such as "some" or "all") can be used to convey not (or not just) their literal meaning but some meaning inferred in context. For instance, it is well known that an "If *P,* then *Q*" statement can be intended and understood not only to convey that *P* is a sufficient condition for *Q* (which corresponds to the literal sense of "if") but also to convey that *P* is a necessary and sufficient condition for *Q* (corresponding to "if and only if") or that *P* is a necessary condition for *Q* (corresponding to "only if"). When interlocutors use "if" in such nonliteral ways, they are not violating any rule of logic or of semantics. They are just making a normal use of language. Logical connectives, it turns out, can be used in the same way as discourse markers such as "therefore" and "but" to suggest implications that should be derived in the context even though they are not entailed by the literal meaning of the utterance.

It might have been tempting to view verbal expressions of logical relationships as similar to verbal expression of arithmetic relationships (as in "A hundred and thirty-five euros divided by five equals twenty-seven euros"). Whether they are done with written digits and special symbols or with words, arithmetic operations obey strict rules of construction and interpretation. Even when they are performed verbally, arithmetic operations are not interpreted on the basis of pragmatic consideration; number words and words such as "divided" and "equal" are used literally.

Unlike verbal arithmetic, which uses words to pursue its own business according to its own rules, argumentation is not logical business borrowing verbal tools; it fits seamlessly in the fabric of ordinary verbal exchanges. In no way does it depart from usual expressive and interpretive linguistic practices.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Statements with logical connectives (or other logical devices), and even sequences of such statements that more or less correspond to syllogisms, are just part of normal language use. They are used by speakers to convey a meaning that cannot be just decoded but that is intended to be pragmatically interpreted. Not only the words used but also the force with which premises and conclusions are being put forward are open to interpretation. They may be intended as categorical or as tentative assertions, hedged by an implicit "in normal conditions."

When you argue, you do not stop using language in the normal way, nor does your audience refrain from interpreting your statements using the same pragmatic capacities they use all the time. In argumentation, ordinary forms of expression and interpretation are not overridden by alleged "rules of reasoning" that might be compared to rules of arithmetic. Rules of arithmetic are taught and are not contested. There is no agreement, on the other hand, on the content and very existence of rules of reasoning. What is sometimes taught as rules of reasoning is either elementary logic or questionable advice for would-be good thinking or good argumentation (such as lists of fallacies to avoid, which are themselves fallacious).[11]

We have been focusing on argumentation, but what about reasoning done "in one's head"? Such individual reasoning, we would argue, is a normal use of silent inner speech.[12] The pragmatics of inner speech have not been properly studied. Given, however, that much inner speech rehearses or anticipates conversations with others, the pragmatics involved are probably not that different from the pragmatics of public speech. When we speak to ourselves, we use words loosely or metaphorically as often as we do in public speech. Our assertions are just as likely to be hedged or qualified with an implicit "probably" or "in normal conditions." There is no reason to assume that when we reason in our head, we follow logical rules that we typically ignore in public argumentation.

Couldn't all the evidence showing that ordinary reasoning isn't governed by the rules of classical logic suggest that it is governed by another kind of logic or by a probabilistic system of inference? Classical deductive logic is "monotonic." This means that if some conclusion logically follows from some initial set of premises, it also follows from any larger set of premises that includes the initial set. As the example of Mary having an essay to write well

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

illustrates, ordinary human reasoning is not monotonic. Not only in daily life but also in scientific, technical, medical, or legal reasoning, conclusions are typically tentative. They may be revised or retracted in the light of new considerations (such as Mary showing up in time for dinner after all). In fact, monotonic inference is, at best, of very limited use for a real-life cognitive system.

In early experimental psychology of reasoning, the nonmonotonic character of ordinary reasoning had been largely ignored or idealized away. In many recent approaches to reasoning, on the contrary, it has been given a central role. Some scholars, such as the cognitive scientist Keith Stenning and the logician Michiel van Lambalgen, aim to replace classical logic with a "nonmonotonic logic" that would provide better insight into the way people actually reason. Others, such as psychologists Mike Oaksford and Nick Chater, argue that reasoning is best viewed not as a logical but as a probabilistic—and more specifically Bayesian—form of thinking.[13]

The project of replacing standard logic with nonmonotonic logic or of replacing logic altogether with probabilities shares a basic presupposition with the traditional approach: that the study of inference must be based on a general and formal understanding of norms of good inference. We are not convinced. We have argued for an evolutionary and modularist view of inferential processes. Every inferential module aims at providing a specific kind of cognitive benefit, and at doing so in a cost-effective way. In this perspective, investigating a given module is a matter of relating its particular procedures to its particular function. The function of the reason module, in particular, is much more specific than that of optimizing knowledge and decision making in general. Investigating general norms of inference, while interesting from a philosophical or from a machine intelligence point of view, may not tell us that much about any specific inference module. It may tell us little, in particular, about the reason module (just as a general theory of locomotion would be of limited use in understanding how bats fly or snakes crawl).

While we doubt that recent nonmonotonic or probabilistic approaches provide the key to understanding reasoning proper, we of course agree with the critics of classical logic that it fails to provide a plausible norm for human inference. Shouldn't we then also agree that classical logic is altogether irrelevant to the study of reasoning? Well, no. Let us venture a limited defense of

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

classical logic, one that defenders of classical logic may not be too happy about. Logic can be used not just as a norm or as a procedure but also as a heuristic tool that clarifies questions and suggests answers. This, we claim, is a main role that logic plays in reasoning. Of course, this goes against the standard view that the function of logic is precisely to overcome the limitations of heuristic thinking.

If syllogisms can be interpreted with some freedom and according to the context, if, even when they are logically sound, they don't necessarily compel rational reasoners to accept their conclusion, what good do they do? What is the point of using them at all? The answer we want to suggest is that they often highlight reasons to accept a conclusion that is not immediately intuitive or to reject a conclusion that is. The very schematism of syllogisms (and deductive relationships generally) tends to exaggerate the degree of logical dependence among our assumptions. In particular, it dramatizes mere incoherencies (where two or more ideas might each give reasons to reject the others) as straight inconsistencies or logical contradictions. Just as exaggerating contours in a picture helps recognition and just as leaving out details in a narrative helps one follow the story, leaving out hedges and ignoring exceptions help focus on reasons that may lead to adopting or rejecting a conclusion. Incoherencies are often hard to detect and to reflect upon. Dressing them up as logical inconsistencies makes them salient targets for the reason module.[14]

Reasoning itself, as we have described it, involves higher-order intuitions about how lower-order intuitions may support some conclusion. True, in principle, the higher-order intuitions might be just about logically relevant properties of reasons and conclusions and might ignore other aspects of their content, but why should they be? This, anyhow, is not what happens. When reasoning about a given issue, higher-order intuitions are about various properties of lower-order intuitions, whether "logical" or not, that are relevant to their value as reasons. Higher-order intuitions in reasoning are metacognitive rather than just "metalogical." Reasoning is based on rich and varied intuitions about intuitions.

Take the kind of reasoning that would classically be represented by means of a disjunctive syllogism with a main premise of the form "*P* or *Q*." Depending, so to speak, on the *P* and the *Q* involved—depending, that is, on content properties that from a standard logicist point of view shouldn't be

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

taken into account—higher-order intuitions may be developed in different directions.

You are reasoning about the location of a book you need and have the intuition that the book must be either in the study or in the living room. You have been searching for it in the study, but so far without success. You decide to look for it in the living room. Of course, if asked to explain why you are now moving to the living room, you might articulate your answer in syllogistic form: "I thought that the book is either in the study or in the living room. It is not in the study; hence it must be in the living room." This would make it sound as if all that was involved was a bit of logic. In fact, your reasoning is more attuned to the particulars of the case. In your mind, you are not ruling out the possibility that you might have missed the book in the study; it is just that the more you looked and failed to find it, the more this became an intuitive reason to conclude that it must be in the living room (or else that you were mistaken in your initial intuition that it had to be in one of these two rooms). The way your reasoning proceeds is sensitive to the fact that a book cannot be in the two rooms at the same time, to the fact that in both rooms there is so much clutter that you might miss it, and to the fact that your initial intuition about the possible location of the book is based on less-than-certain memory. Even when you express it as a syllogism, you expect your audience to take this as just a schematic rendering of your thinking and to use their own richer intuitions to understand and evaluate your reasons.

Take now a different example, where the schematism is the same but the relevant intuitions quite different. Seeing Molly frown, you intuit that she must be upset about something that occurred yesterday or worried about something that might occur. Which is it? She tells you that what occurred yesterday didn't upset her and you assume that she is sincere. You conclude that she is probably worried. Your initial intuition that Molly might be upset or worried didn't quite exclude that she might be both. Her sincerely saying that she was not upset didn't rule out that she might be upset without realizing it. Your intuitive reasons favor, then, a somewhat tentative conclusion: you tend to believe that she is worried rather than upset, or at least more worried than upset. Suppose now that you tell Ramon, a common friend, "Molly is worried!" and he answers, "I think she is just upset about what happened yesterday." You might then argue in syllogistic form: "Looking at Molly's face, you cannot tell whether

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

she is upset or worried; she says, however, that she is not upset, not about yesterday, not about anything, and she is manifestly sincere; hence she must be worried." Again, this would schematize your reasoning, but in so doing, it would present Ramon with a clear challenge and with reasons to revise his beliefs and either to conclude just that Molly is worried or to view things in a more nuanced way, as you do yourself.

In argumentative reasoning in particular, the use of logical relationship plays a heuristic role for one's audience. It helps challenge them to examine and enrich or revise their beliefs or else to defend them with arguments in their turn. Thanks in part to its logical garb, argumentation, if not always convincing, is at least quite generally challenging.

More generally, however, reasoning is not the use of logic (or of any similar formal system) to derive conclusions. But what, then, is the method of reasoning (if there is one)?

Is There a Method for Reasoning?

Reasoning, as we have described it so far, is rather limited. Humans reason when they are trying to convince others or when others are trying to convince them. Solitary reasoning occurs, it seems, in anticipation or rehashing of discussions with others and perhaps also when one finds oneself holding incompatible ideas and engages in a kind of discussion with oneself. Just as with justifications, the production of arguments proceeds by means of backward inference, from a favored conclusion to reasons that would support it.

Surely, something must be missing in this picture: one is often prompted to reason not by a clash of ideas (with others or within oneself) but by self-addressed questions. Moreover, isn't the point of such individual, inquisitive reasoning to discover the right answer to a question of one's own rather than to confirm an already favored conclusion?

There are indeed questions that we can and do approach on our own, without any bias, any hunch in favor of this or that answer, and that we are able to answer by reasoning forward, from premises to conclusion. These, however, tend to be special kinds of questions that are approached with ad hoc reasoning methods rather than with everybody's everyday reasoning dis-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6						8	
			4	1	9			5
				8			7	9

Figure 15. A Sudoku grid.

positions. The simplest examples of such questions are found in games or puzzles devised to entertain, teach, or test people.

Take the game of Sudoku (see the example in Figure 15). It is played on a square grid of eighty-one cells, some already containing a digit at the beginning of the game, the others blank.

The task is to find the digit that each of the blank cells should contain, knowing that every digit between 1 and 9 must occur once and only once in every vertical column, every horizontal row, and every three-by-three box indicated by thick lines.

Sudoku players approach the task in an impartial manner. They have no a priori hunch and no stake whatsoever in any particular solution. They know that in every grid there is one and only one right solution for each and every cell, and the players' goal is to find them all. That much is clear. What is less clear is how the players proceed.

In the simplest cases, it is possible to find the digit that goes in a given cell by means of a simple elimination method until only one possibility is left. Which digit, for instance, should go in the central (grayed in the figure) cell?

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Looking at the grid, you find reasons to eliminate digit 1 (it already occurs in both the horizontal and the vertical rows that contain the grayed square), digits 2 and 3 (they already occur in the three-by-three box), and so on with all nine digits except 5, which therefore must be the digit that goes in this central cell. Sudoku is a perfect illustration of Sherlock Holmes's famous maxim, "When you have eliminated the impossible, whatever remains, however improbable, must be the truth" (and of the very limited usefulness of "thinking like Sherlock Holmes").

The elimination method we just illustrated may help players find a few missing digits, but it will not help them fill the whole grid. To fully solve a Sudoku puzzle, even a simple one, players must use more complex methods. The real challenge for psychologists, however, is not so much to understand how players apply methods they have been taught but how some of them at least discover useful methods on their own.[15] After all, the reasoning involved in discovering these methods is much more impressive than that involved in solving a puzzle once you know what method to apply.

How do ordinary Sudoku players discover new methods? We argued that in general, when people reason, they start from an intuitive bias or a hunch in favor of a given conclusion and look for reasons in its favor through backward inference. While this is not the way players *use* methods to solve a Sudoku puzzle, it is, we suggest, the way players *discover* these very methods in the first place (when they don't learn them from others). Merely understanding the rules of Sudoku makes the use of the simple elimination method we have described fairly intuitive. Practice familiarizes players with various subtle regularities of Sudoku grids and provides them with intuitive insight into more and more elaborate methods for which successful applications provide confirming reasons.

The discovery of effective Sudoku methods may well be, then, yet another instance of basic biased reasoning: a search for reasons in favor of an initial intuition—not, in this case, an intuition about the digit that goes in a given cell, but a higher-order intuition about the kind of considerations that may allow the player to narrow down the range of possible digits for every cell. If we are right, this means that people do not use a general higher-order explicit method to discover the various, more specific, explicit methods needed to solve a Sudoku puzzle.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Sudoku puzzles are problems that, with practice and explicit methods, people can solve. Does this generalize to reasoning problems across domains? Can one, with the right method, effectively answer questions of any kind? The short answer is no.

Humans are, it is true, capable of applying methods that they have been taught or that they have discovered on their own to do a great variety of things: solve a Sudoku puzzle, construct a Lego castle, find a word in an alphabetic dictionary, buy from a vending machine, bake a cake, convert Roman into Arabic numerals, learn the basic moves of tap dancing, find solutions to arithmetic or geometric problems, or use databases to answer queries about, say, legal precedents, life expectancy, or the yield of equities. This ability to understand and apply step-by-step methods is a hugely important aspect of human psychology. It plays a major role in the development and transmission of cultural skills, including specialized problem-solving skills. The psychologist Keith Stanovich has argued that the ease with which people acquire and apply such methods (or "mindware") correlates with individual differences in rationality.[16] Still, reasoning doesn't consist in applying such methods, and in general, it doesn't need them. Applying these methods is no more quintessential reasoning than military marching is quintessential walking.

Most of the questions that people encounter in their daily life or in the pursuit of their longer-term interests cannot, in any case, be answered by following instructions. Every year, new psychological counseling books offer to instruct you on how to reason better in business, love, friendship, and games, promising to do much better than last year's books that made the same promises. While they may at times include some sensible advice, all these books come short of delivering on their promises. Achieving such desirable results involves understanding a great many aspects of the world around us, and knowing which to take into account and which to ignore in any given situation; it involves identifying and tackling an endless variety of issues along the way; there are no adequate instructions for reasoning effectively about most real life problems.

What differentiates Sudoku puzzles from most issues on which we might reason is that they are perfectly well-defined. The wider context is not relevant, the evidence is all here, there is one and only one correct solution, and any incorrect solution leads to clear inconsistencies. In psychology, such

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

problems with a definite and recognizable solution and some effective methods to discover it have often been studied under the label of "problem solving" rather than "reasoning." Solving such problems when the right method is not known typically involves some reasoning but also trial-and-error tinkering and task-specific forms of insight.

Many games and many serious problems in mathematics, logic, programming, or technology involve problem solving in this sense. Scientist, technicians, and laypersons have discovered and developed methods to address problems that can be solved in such a way. Some of these methods have become cultural success stories because of their practical applications, scientific relevance, intellectual elegance, or appeal as leisure activities.

To psychologists of reasoning, problems and puzzles that can be solved in clear and effective ways, especially problems devised on purpose by experimentalists such as the Wason selection task, may seem to provide optimal material for experimental investigation. The temptation has been to assume not only that some of these problems are easy and fun to experiment with but also that they provide crucial evidence on the basic mechanisms of reasoning. Such narrowly circumscribed tasks are, however, special cases among the quite diverse challenges that reasoning helps us confront. Problem solving involves some reasoning, but most reasoning doesn't even resemble problem solving in the narrow sense in which psychologists use the phrase.

When people reason on moral, social, political, or philosophical questions, they rarely if ever come to universal agreement. They may each think that there is a true answer to these general questions, an answer that every competent reasoner should recognize. They may think that people who disagree with them are, if not in bad faith, then irrational. But how rational is it to think that only you and the people who agree with you are rational? Much more plausible is the conclusion that reasoning, however good, however rational, does not reliably secure convergence of ideas.

Scientists, it is true, do achieve some serious and, over time, increasing degree of convergence. This may be due in part to the many carefully developed methods that play a major role in conducting and evaluating scientific research. There are no instructions, however, for making discoveries or for achieving breakthroughs. Judging from scientists' own accounts, major ad-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

vances result from hunches that are then developed and fine-tuned in the process of searching, through backward inference, for confirming arguments and evidence while fending off and undermining counterarguments from competitors.

Arguably, science is the area of human endeavor where rationality and good reasoning are most valued (as we will discuss in detail in Chapter 18). Reasoning, however, does not cause scientists to spontaneously converge on the best theories. It causes them, rather, to elaborate and vigorously defend mutually incompatible competing theories. It also helps them—with a higher, even if still imperfect, degree of convergence this time—to evaluate competing theories and reach some degree of tentative agreement on the current winners of the competition. Winning in this competition may well be, to a large extent, a matter of epistemic luck—having invested in a better initial hunch for which there are stronger evidence and better arguments to be found—more than of better reasoning in producing new ideas. What scientific practice suggests is, on the one hand, that various specific methods can be relevant to specific reasoning tasks and, on the other hand, that there is no such thing as a general method for reasoning.

The study of reasoning has been dominated by a normative goal and a descriptive expectation: the goal of discovering and making explicit a general method that could produce good reasoning and the expectation that actual human reasoning would be guided by an approximation of such a method. Both the goal and the expectation have, so far, been disappointed. This failure is not an accident. There is a principled explanation for it. The procedures of intuitive inference, we have argued, are unconscious, opportunistic, and diverse. The idea that intuition might consist in following a general method makes little sense.

Our higher-order intuitions about reasons for some intuitive conclusion take into account many of the properties of the lower-order intuitions they are about, properties that may vary greatly from one domain to another. This is not bad, let alone fallacious, reasoning. In putting our reasons in argumentative form to convince others, sometimes we appeal to their own nuanced higher-order intuitions, sometimes we choose to be schematic, and often we do a bit of both. The schematism we employ, rooted in ordinary language, has

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

inspired one of the great intellectual achievements of human history: the development of the science of logic. Logic, however, tells us neither how we reason nor how we should reason. There is no general method that we could and should follow when reasoning, either on our own or in dialogue with others.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

10

Reason: What Is It For?

Look at the piece of furniture that appears in Figure 16 on the following page. Hardly a well-designed chair, is it? The seat is so low that you'd practically have to squat on it. The back is too high; the upper bar is a horizontal slab on which you could not lean comfortably. How come? Actually, this is not a chair. It is a church kneeler. The height is right for kneeling. The "back" is in fact the front, and the upper slab is there to rest not one's back but one's praying hands. Once this artifact's real function is recognized, what looked like flaws turn out to be well-designed features.

Like the object in Figure 16, reason seems to have an obvious function: to help individuals achieve greater knowledge and make better decisions on their own. After all, if using reason doesn't help one reach better beliefs and choices, what is it good for? However, like a kneeler used as a chair, reasoning serves this function very poorly.

We've already examined quite a few failures of reasoning, and we will look at many more in Part IV, but the upshot of experimental research on how we reason is that we do so in a way that is biased and lazy and that often fails to solve even simple but unfamiliar problems. It would be easy to stop here and simply conclude, as many psychologists have done, that human reason is just poorly designed.

Another conclusion is possible. The true function of reasoning may have been misunderstood. Reasoning might be the kneeler-mistaken-for-a-chair of modern psychology. In Chapters 7, 8, and 9, we have argued that reasons are commonly used in the pursuit of social interaction goals, in particular to justify oneself and to convince others. Here, we adopt an evolutionary approach

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

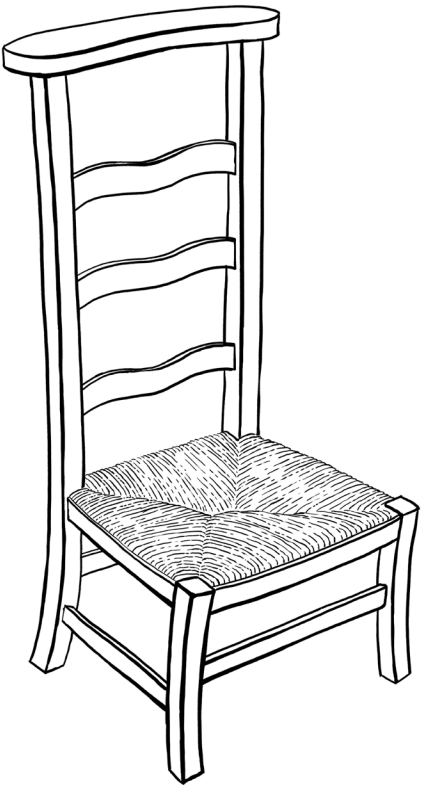

Figure 16. A poorly designed chair?

to argue that these social uses of reason are not side effects or minor functions of a cognitive mechanism having as its main function the enhancement of individual cognition and that, on the contrary, the main function of reason is social. Why resort to an evolutionary approach? Because this is the only approach that explains the fact that complex inheritable traits of living things tend to produce beneficial effects. Outside of an evolutionary perspective, it is quite unclear why human reason, or anything else for that matter, should have any function at all.

Functions before and after Darwin

That human reason has a function is an idea commonly taken for granted and one that we want not to challenge but to develop on a new basis. To begin

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

with, let us be clear about the very notion of a function: discussions about the evolution of reason are often marred by a superficial understanding of the notion.

What is a function? How can one decide what is the function of an artifact such as a chair, a natural quality such as a flower's scent, a body part such as a wing, or a mental capacity such as human reason? The two questions—what is a function, and how do we identify a particular function—are closely related. For artifacts, a rough answer is generally easy to produce. The function of an artifact is what it is made for. Who decides what an artifact is for? Whoever designed it. A chair is made for sitting, a kneeler for kneeling, and so on.

Attributing a function to a biological trait is much less straightforward. No one made wings for flying or the heart for pumping blood. And yet, who would doubt that wings are for flying? Who would contest William Harvey's discovery of the function of the heart?

Before Darwin, attributing a function to some trait of an animal or a plant involved little more than answering the question: What is it good for? Find a good effect of a biological trait and just assume that what it is good for is what it is for. What could be more useful than flying? So, the function of wings is to enable flying. Such a commonsense notion of function is not useless; it has played an important role in science before Darwin, such as in Harvey's discovery of the function of the heart, and it still plays at least a heuristic role in modern science. What question such an ahistorical notion of function raises and fails to solve is: Why, in the first place, should we expect biological traits to have useful effects? We don't expect anything of the sort in the case of other natural objects such as stones, planets, and subatomic particles. These don't have functions. For a long time, the only answer was that God, the maker of all things, has made wings for flying, hearts for pumping blood, and so forth—an answer that replaced a series of interesting problems with an unfathomable mystery.

It was Darwin's theory of natural selection that provided the basis for a coherent and useful notion of a function based on a genuine explanation of why organs and other traits of living things have the functions they have.

Any inheritable trait of an organism has many effects. Most of these effects are without consequences for reproductive success. The scent of a flower, for instance, might make some local animals dizzy without this harming or

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

benefiting the plant in any way. Some effects of a trait may happen to enhance reproductive success: the scent of a flower may attract insects and recruit them as pollinators. Still other effects of a trait may actually compromise reproductive success: a scent might attract florivores and entice them to eat the flower, thereby diminishing the amount of pollen available to pollinators.

When overall the effects of a heritable trait are more beneficial than harmful to the reproductive success of its carriers, then this trait is likely to be selected and to propagate over generations. Selected traits have at least one of their effects that has contributed to the reproductive success of their carriers and thereby to their own propagation. The scent of wildflowers, for instance, has been selected because by attracting pollinators, it contributed to the reproductive success of the plants. When biologists talk of the function of a selected trait, they refer to such a beneficial effect.

Selected traits may have more than one function, and these functions may change in the course of evolution. Penguins, for instance, are descendants of birds that had used their wings to fly, but the function of penguins' wings is to help them swim. At some point in their evolution, these wings must have served both functions, as they do in some seabirds, such as in razorbills.

Traits that owe their propagation to their function(s) are *adaptations.* Not all features of an organism are adaptations. Some features result from physical or chemical constraints on biological development. The fact that all organisms contain carbon, for instance, is a basic chemical property of life itself (as it is found on earth, at least). Not all features of an adaptation are adaptive, either. Many are side effects. The regular contractions of the heart are an adaptation. They have the function of causing blood flow throughout the body. The noise produced by these contractions, on the other hand, isn't adaptive. It is a side effect.

The Function of Reason from Aristotle to Twentieth-Century Psychologists

Just like ideas about the function of wings, ideas about the function of reason emerged well before Darwin and the development of a scientific notion of function. Even after Darwin, much evolutionary thinking on reason shares with pre-Darwinian views a sense that the function of reason, just as that of

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

wings, is obvious enough. Reason is a means for individuals to acquire superior knowledge and to make better decisions. Reason, by performing this intellectual function, elevates humans above all other animals.

Darwin himself expressed this same general idea: "Of all the faculties of the human mind," he wrote in *The Descent of Man,* "it will, I presume, be admitted that *Reason* stands at the summit."[1] He put, however, an evolutionary twist on this classical view. Through his intellectual faculties, he argued, man "has great power of adapting his habits to new conditions of life. He invents weapons, tools, and various stratagems to procure food and to defend himself. When he migrates into a colder climate he uses clothes, builds sheds, and makes fires; and by the aid of fire cooks food otherwise indigestible." Because these intellectual faculties are variable and tend to be inherited, "they would have been perfected or advanced through natural selection. . . . In the rudest state of society, the individuals who were the most sagacious . . . would rear the greatest number of offspring."[2]

The functional effects of reason are roughly the same for Darwin as they were for Aristotle. What is new with Darwin, however, is the use of these effects to explain why reason should have evolved. And because Darwin is making a more precise claim about the function of reason, his claim is more open to a scientific challenge. If reason evolved to help individuals think on their own, then it should really provide for truly better thinking in terms of both cognitive benefits and mental costs, and if it does not, as much of modern psychology of reasoning suggests is the case, then we are indeed faced with a serious problem.

Psychologists who discovered what looked to them like major flaws of human reason must have, you would think, understood and discussed the challenge this presented for a post-Darwinian understanding of the function of reason. Actually, they did not. Until the 1990s, evolution, natural selection, and biological function were hardly ever mentioned in the psychological literature on reasoning. Psychologists just took for granted that the function of reason—with "function" understood as a commonsense rather than biological notion—was to enhance individual cognition. They then concluded that reason was not performing its function as well as had been assumed.

If anything, psychologists saw flaws in reasoning as evidence against an evolutionary approach. Evolved mechanisms should perform their function well

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

enough to have been selected. Reasoning, with all its flaws, is not properly geared to the pursuit of knowledge and good decision. Hence, many concluded, the case of reasoning shows the irrelevance of the evolutionary approach to human psychology.[3] Of course, psychologists might have rethought the function of reason instead of uncritically accepting common wisdom in the matter, but they did not.

While the foundations of an evolutionary approach to psychology were laid by Darwin himself and some relevant work had been done over the years, especially in the study of animal behavior, a general and systematic discussion of the evolution of human cognitive mechanisms and of reasoning in particular started only when psychologist Leda Cosmides and anthropologist John Tooby outlined an ambitious program of "evolutionary psychology" in the 1980s.[4]

"Has Natural Selection Shaped How Humans Reason?"

From the start, human reasoning has been a major topic of research in evolutionary psychology and a main focus of polemics for scholars opposed to this approach. The polemic erupted when, in 1989, Leda Cosmides published what became the most controversial article in the history of psychology of reasoning. It was entitled "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason?"[5] (hence the title of this section).

There were in Cosmides's article two major theoretical claims. From an evolutionary point of view, she argued, one should expect reasoning mechanisms to have evolved as responses to specific problems that had been recurrent in the ancestral environment. Specialized adaptations do a much better job of addressing the problems they evolved to handle than would any general problem-solving ability. In fact, it is not even clear what an ability to solve problems in general might consist of (if not, precisely, of a complex articulation of a great many more specialized mechanisms). A general reasoning ability, if such a thing could evolve at all, might, at best, help in dealing with the residue of problems not properly handled by specialized mechanisms. On tasks of no evolutionary significance (such as those typically given in reasoning experiments), Cosmides argued, one should expect reasoning to be much less effective than on problems for which dedicated mechanisms have evolved.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Cosmides's second and central claim was that among the many cognitive adaptations most likely to have evolved, there had to be an inferential mechanism aimed at solving a major problem raised by cooperation. Cooperation is an interaction between two or more individuals where each incurs a cost and receives a benefit. Each member of a party of big game hunters, for instance, spends time and takes risks to help the group catch a prey. Provided that the benefits are greater than the costs and are shared fairly, cooperation is advantageous to all cooperators. However, in many cases, each cooperator stands to gain even more by sharing in the benefits of cooperation without paying the full cost, in other terms by cheating or free riding. Hunters who avoid taking risks but take the same share of the meat as others are at an advantage. Of course, if none of the hunters take risks, the prey will escape. Widespread free riding is likely to result in failure of cooperation. For cooperation to be profitable and hence to endure, cheaters must be identified and either controlled or excluded.

Cosmides and Tooby argued that the identification of cheaters should be seen as a major problem in human evolution, favoring the emergence of a specialized module capable of computing the rights and duties of cooperators as a basis for detecting cheaters. To test this hypothesis, they relied massively on the Wason selection task, which we talked about in Chapter 2 and which we don't see as adequate for the job.[6] Our goal here, however, is not to evaluate Cosmides's arguments for the existence of an evolved ability to detect cheaters (which we see as strong) or her experimental evidence (which we see as weak). It is to use her approach as a source of inspiration and of contrast to propose another, no less radical and even more ambitious evolutionary hypothesis about the function of human reason in general, and reasoning in particular.

Like Cosmides and Tooby, we expect an evolved mind to consist principally in an articulation of modular mechanisms. Modules are specialized—each helps solve some specific type of problem or take advantage of some specific type of opportunity in a way that contributes to biological fitness. Reasoning as classically understood, on the other hand, is allegedly in the business of addressing any kind of problem and finding ways of taking advantage of any kind of opportunity. Classical reasoning is not a specialist but a generalist, not a narrow module but a broad faculty.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

If the job of reasoning is much too broad to be that of a module, there seem to be only two possible conclusions: either the central and most distinctive component of the human mind is not modular, or else reasoning as classically understood doesn't even exist. Cosmides and Tooby favor the second conclusion: contrary to a long-held dogma of philosophy and psychology, there is no such thing as reasoning in general.[7] Other researchers working on the evolution of the mind, such as the psychologist Cecilia Heyes and the philosopher Kim Sterelny, favor the first possible conclusion: the existence of domain-general reasoning is evidence that the human mind is much less modular than Cosmides and Tooby and others have argued.[8]

Are these the only two possibilities? No. As we have shown (in Chapter 6), a metarepresentational module processes a very special type of object, representations, and attends to properties specific to these representations. Such a module may nevertheless indirectly provide information about the very topics of the representations it processes. Mindreading, for instance, informs us about not only the thoughts of others but also all the things and events these thoughts are about. Reasoning, we have argued, is produced by a metarepresentational module, the specific domain of which is the relationship between reasons and the conclusions they support. These reasons and conclusions, however, can themselves be about any topics. As a result, inferences about reasons-conclusions relationships indirectly yield conclusions in all domains, indirectly providing a kind of virtual domain-generality.

So, to answer the question posed in the subtitle of Cosmides's article and in the title of this section, we agree that natural selection has shaped how humans draw all kinds of inferences and has produced a wide variety of specialized inferential modules. One of these, we add, is a reason module. The justifications and the arguments that the reason module produces may have, embedded in them, conclusions relevant to all domains of knowledge and action. This virtual domain-generality does not, however, make reason, and the organization of human mind generally, any less modular.

What functions does the reason module fulfill? We have rejected the intellectualist view that reason evolved to help individuals draw better inferences, acquire greater knowledge, and make better decisions. We favor an interactionist approach to reason. Reason, we will argue, evolved as a response to problems

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

encountered in social interaction rather than in solitary thinking. Reason fulfills two main functions. One function helps solve a major problem of coordination by producing justifications. The other function helps solve a major problem of communication by producing arguments. (Our earlier work has been focused on reasoning and on developing, within the interactionist approach, an "argumentative theory of reasoning.")[9]

The Challenge of Coordination and the Justificatory Function of Reason

Human cooperation is exceptional not only by its scale but also by the open-ended variety of the forms it takes. Other animals may have a few types of cooperative interactions in their behavioral repertoire with little or no place for creative improvisation. What each cooperator may expect of the others is largely predetermined. When these expectations are not met because of the incompetence or the defection of one of the cooperators, cooperation is likely to fail. Humans, on the other hand, rely less on predefined expectations. Existing forms of cooperation are deployed with great flexibility and often readjusted on the fly. New forms are often tested. This flexibility and creativity of human cooperation can be highly advantageous, but only if massive cognitive resources are invested to secure effective coordination.

To achieve the degree of fine-grained coordination that their multiple forms of cooperation require, humans need mutual expectations that have to be constantly updated to remain reliable. Members of a party of warriors, or of a sports team, for instance, readjust or even redefine their tactics and their mutual roles on each occasion. They must, moreover, be able to rely on each other not only when things go as anticipated but also when they don't. So must, in different ways, friends, spouses, coworkers, and business partners.

Even among competent cooperators, there may well be differences in the understanding of the common goals and of the part each individual must play. Differences of interests, instead of leading to defection or cheating, are often recognized and handled on the basis of mutual commitment to what is seen as fair; here too, however, there may be differences of interpretation leading to failures of mutual expectations.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

For humans, knowing what to expect of each other is a crucial cognitive challenge. How is this challenge met? How do humans succeed in forming, if not perfect, at least adequate mutual expectations? The most common answer consists in invoking two mechanisms: norms at the sociological level, and understanding of the mental states of others at the psychological level.

At the level of the social group, there are shared norms of various kinds: moral, legal, religious, prudential, technical, and so on. They may be explicitly codified or not and enforced with sanctions or not. These norms regulate a great variety of social interactions.[10]

Some norms aim directly at securing coordination in a way that is beneficial to all the people involved. Without traffic rules, for example, driving a car would be absurdly risky. Very precise rules of coordination, however, tend to be highly specialized. The kind of interaction traffic rules regulate, for instance, is unlike any other; at every moment, drivers have a relatively narrow, well-defined range of options. If they don't coordinate their decisions, their lives are at risk. Effective coordination is in the interest of all of them.

Legal, moral, or religious norms, on the other hand, contribute to all kinds of social coordination in a great many ways, but securing effective coordination is not their sole purpose and needn't be their main one. These norms leave much room for interpretation and many relevant issues untouched. In some milieus, for instance, not arriving late at a dinner party may be a faux pas. People, moreover, don't at all times have the same interest in these various norms' being obeyed. Elected officials coming up for reelection, for instance, overtly encourage all their constituents to follow the norm and vote while making it harder to do so for those less likely to reelect them.

Do the many and diverse norms found in one's society constrain people's options to the point of indicating what people can expect from each other in the fine-grained manner needed for effective coordination? Of course not! There are, for instance, many norms—legal, moral, religious—telling spouses what they can expect of each other. Still, to achieve the level of mutual expectations that daily life coordination requires, spouses need to understand much more about each other than the fact that they are supposed to abide by a socially sanctioned set of norms.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

In many forms of social interaction, some degree of creativity and improvisation is socially condoned or even required. You throw a party, for example. Whom should you invite? Conventional or moral norms might suggest that you invite people who have invited you in the past, so perhaps you should invite Olga, who invited you to her party last month. For other possible guests, norms of reciprocity won't help at all, but there are still issues of coordination to be addressed. If you invite Diego, for instance, you had better not invite Ruth. If you invite Ruth, she will want you to invite Chao.

Your future interactions with others will depend on many small, interconnected decisions you make. So what should guide you in solving coordination problems that norms leave unanswered? Here the standard answer is: mindreading. You have to understand the states of mind of Olga, Diego, Ruth, and other possible guests in order to anticipate their reactions and secure the conditions for a successful party that will enhance rather than compromise your future relationships with others.

Just like norms, mindreading plays an essential role in coordination. Still, the picture is far from complete. What is blatantly missing is the fact that individuals don't just infer what they can expect from each other on the basis of what they know of other people's minds and of the norms they share. People's mutual expectations are reviewed, discussed, negotiated in detail. Many decisions on how to interact are themselves taken interactively.

Gossip provides rich evidence of what can be expected of third parties. Individuals, however, need not be passive objects of gossip. They can participate in the ongoing conversation about themselves, explain and justify their views and their decisions, and, in so doing, to some extent, safeguard their own reputation. Officials trying to suppress the vote of constituents unlikely to reelect them justify their actions by saying they are just trying to prevent voting fraud. Guests arriving too late at a party justify themselves by explaining how they were delayed. You can tell Diego that, alas, you had to invite Ruth because she had invited you before; you can tell Ruth that, of course, she can come with Chao because her friends are your friends; and so on.

By giving reasons to explain and justify yourself, you do several things. You influence the way people read your mind, judge your behavior, and speak of you. You commit yourself by implicitly acknowledging the normative force

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

of the reasons you invoke: you encourage others to expect your future behavior to be guided by similar reasons (and to hold you accountable if it is not). You also indicate that you are likely to evaluate the behavior of others by reasons similar to those you invoke to justify yourself. Finally, you engage in a conversation where others may accept your justifications, question them, and invoke reasons of their own, a conversation that should help you coordinate with them and from which shared norms actually may progressively emerge.

Reducing the mechanisms of social coordination to norm abiding, mindreading, or a combination of these two mechanisms misses how much of human interaction aims at justifying oneself, evaluating the reasons of others (either those they give or those we attribute to them), criticizing past or current interactions, and anticipating future ones. In these interactions about interactions, reasons are central.[11]

Justificatory reasons, in fact, bridge the gap between norms and mindreading. When we justify ourselves, we present our motivations as normatively apt, and we present norms as having motivating force. In other terms, we psychologize norms and "normalize" mental states. In doing so, our goal is not to give an objective sociological or psychological account of our actions and interactions; it is to achieve beneficial coordination by protecting and enhancing our reputation and influencing the reputation of others.

The role of reasons in social coordination has often been highlighted in philosophy, psychology, and the social sciences. The dominant view, however, is that attributing reasons is the most elaborate form of mindreading. We would argue that as far as mindreading goes, the attribution of reasons is typically misleading. The causal role it gives to reasons is largely fictitious; the reasons people attribute to themselves or to others are chosen less for their psychological accuracy than for their high or low value as justifications. The explanatory use of reasons, we suggest, is in the service of its justificatory use: it links reasons to persons so that good reasons are seen as justifying not just a thought or an action but also the thinker of that thought, the agent of that action.

The ability to produce and evaluate reasons has not evolved in order to improve psychological insight but as a tool for defending or criticizing thoughts and actions, for expressing commitments, and for creating mutual expectations. The main function of attributing reasons is to justify oneself and to evaluate the justifications of others.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

The Challenge of Communication

Even more than human cooperation, human communication stands apart by its scale, its diversity, and the complexity of its mechanisms. Other animals use a small repertoire of signals to convey a few basic messages about the here-and-now: warnings about the presence of a predator, threats to competitors, and mating calls, for example. Humans have languages with huge vocabularies and powerful syntax with which they can produce an unbounded variety of linguistic utterances. With the help of language, they are able to communicate simple or complex ideas about any conceivable topic, about events distant in space and time, about general facts, or about abstractions, topics that are absent from animal communication and that are at the center of human cultural knowledge.

Humans, who have by far the richest codes, use them to communicate even more than what they manage to encode. As we have already noted, when speaking, people do not fully encode what they mean. What they do is give partly encoded evidence of their meaning. Hearers infer the speaker's intended meaning from this linguistic evidence taken together with contextual information. When, for instance, Azra tells Marco, "This case is too hard," he understands her to mean that the legal case on which he had chosen to write an essay for the tort law class is too difficult and that he should select an easier case to write about. In so understanding Azra, Marco goes well beyond decoding the sentence she uttered—a sentence that in other contexts could be used to convey utterly different meanings—but he does not go beyond the ordinary pragmatic procedures of comprehension we all use all the time.

Securing comprehension is, however, only half of the goal of communication. A speaker typically wants not only to be understood but also to be believed (or obeyed), to have, in other terms, some influence on her audience. A hearer typically wants not just to understand what the speaker means but, in so doing, to learn something about the world. This occurs when the hearer not only understands what he is told but also accepts it. Azra's intention in telling Marco "This case is too hard" was to have her implicit advice accepted. His goal in paying attention to what she was saying was to gain some guidance from it. Still, Marco may have understood what Azra was telling him and disagreed. In human communication, comprehension does not automatically

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

secure acceptance. (Do you, for instance automatically believe everything you understand when reading a book like this one? Of course not!)

Without acceptance of the information communicated, communication wouldn't be beneficial to either communicators or addressees. If it were not beneficial to both, communication could not evolve. Would, then, a disposition in addressees to automatically accept communicated information ensure optimal communication? Far from it. Imagine what communication would be like if people automatically believed or did what they were told. Communicators might be quite satisfied; their ability to influence gullible and docile listeners would be without limit. Addressees, on the other hand, accepting everything they were told, would be prey to all kinds of misinformation and manipulation, not a satisfactory condition at all.

Automatic trust would make more sense only if there were a corresponding disposition in communicators to be automatically trustworthy. The eighteenth-century Scottish philosopher Thomas Reid affirmed that God had equipped humans with such a pair of dispositions:

> The wise and beneficent Author of nature, who intended that we should be social creatures, and that we should receive the greatest and most important part of our knowledge by the information of others, hath, for these purposes implanted in our natures two principles that tally with each other. The first of these principles is a propensity to speak truth. . . . [The second principle] is a disposition to confide in the veracity of others, and to believe what they tell us.[12]

An omnipotent creator might have decided to make humans both trusting and truthful. Natural selection is much less likely to produce and maintain such perfect harmony. It will, if possible, favor communicators who do not hesitate to deceive trusting receivers when it is in their interest to do so. Likewise, it will favor receivers who are not easily deceived.

Communication can still evolve under natural selection when the biological interests of emitters and receivers are, if not identical, at least sufficiently aligned, such as in the case of honeybees (which in the case of sterile workers contribute to the propagation of their genes by all helping the queen reproduce). Communication can also evolve when it is limited to topics of common

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

interest where the communicator wouldn't gain by deceiving its audience or the audience by distrusting the signal, such as when a female baboon signals to potential mates, by means of a large swelling of her posterior, that she is ready to mate.

Human communication, however, is definitely not limited to topics of common interest where truthfulness and trust are mutually advantageous to the interlocutors. Linguistic signals can be produced at will to inform or to mislead. Unlike the honeybee's waggle dance or the female baboon's sexual swelling, linguistic signals are not intrinsically reliable. Human communication takes place not only among close kin or cooperators but also with competitors and strangers. Lying and deception are in everybody's repertoire. Even children start expecting and practicing some degree of deception around the age of four.[13]

Humans stand to gain immensely from the communication of others. Without communicating a lot, they couldn't even become competent adults or lead a normal human life. As a result of their dependence on communication, humans incur a commensurate risk of being deceived and manipulated. How, notwithstanding this risk, could communication have evolved into such a central, indispensable aspect of human life?

Communication is a special form of cooperation. The evolution of cooperation in general poses, as we saw in Chapter 9, well-known problems. It might seem reasonable to expect that theories that explain the evolution of human cooperation might also explain the evolution of communication.[14] Do they really?

Actually, communication is a very special case. In most standard forms of cooperation, cheating may be advantageous provided one can get away with it. For any given individual, doing fewer house chores, loafing at work, or cheating on taxes, for example, may, if undetected, lower the costs of cooperation without compromising its benefits. This being so, cooperation can evolve and endure only in certain conditions—in particular, when organized surveillance and sanctions make cheating, on average, costly rather than profitable, or when the flow of information in society is such that cheaters put at risk their reputation and future opportunities of cooperation. For communication to evolve, however, the conditions that must be fulfilled differ considerably from this.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Although this is hard to measure and to test, there are reasons to think that communicators are spontaneously honest much of the time, even when they could easily get away with some degree of dishonesty. Why? Because when humans communicate, doing so honestly is, quite commonly, useful or even necessary to achieve their goal. People communicate to coordinate with others, to organize joint action, to ask help from friends who are happy to give it. In such situations, deceiving one's interlocutors, even if undetected, wouldn't be beneficial; it would just be dumb.

Still, in other situations that are also rather common, a degree of deception may be advantageous. People generally try to give a good image of themselves. For this, even people who consider themselves quite honest routinely exaggerate their virtues and achievements, use disingenuous excuses for their failings, make promises they are not sure to keep, and so on. Couples who most of the time may be cooperative and honest with one another lie about their infidelities. Then there are cases where candid communication is generally seen as just incompetent—these include politics, advertisement, propaganda, and some businesses where honesty results in too many missed opportunities. Being lied to may be less frequent than being told the truth, but when it happens, it can be quite costly.

On the one hand, even the most sincere people cannot be trusted to always be truthful. On the other hand, people who have no qualms lying still end up telling the truth rather than lying on most occasions, not out of honesty but because it is in their interest to do so. Hence, standard ways of controlling or excluding cheaters might not work so well in the special case of liars. An ad hoc version of a tit-for-tat strategy[15]—you lie to me, I lie to you; or you lie to me, I stop listening to you—might harm us as much or more than it would punish liars. An ad hoc version of a "partner choice" strategy[16]—you lie to me, I cease listening to you and listen to others—would, if applied systematically, deprive of us of much useful information.

More generally, to benefit as much as possible from communication while minimizing the risk of being deceived requires filtering communicated information in sophisticated ways. Systematically disbelieving people who have been deceitful about some topic, for instance, ignores the fact that they may nevertheless be uniquely well informed and often reliable about other topics. Automatically believing people who have been reliable in the past ignores the

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

fact that, on some new topic where it is in their interest to deceive us, they might well do so. Well-adjusted trust in communicated information must take into account in a fine-grained way not just the past record of the communicator but also the circumstances and contents of communication.

Epistemic Vigilance

Communication is, on the one hand, so advantageous to human beings and, on the other hand, makes them so vulnerable to misinformation that there must have been, we have suggested in earlier work,[17] strong and ongoing pressure for developing a suite of mechanisms for "epistemic vigilance" geared at constantly adjusting trust. There is no failsafe formula to calibrate trust exactly, but there are many relevant factors that may be exploited to try to do so. Reaping the benefits of communication without becoming its victim is so important that any mechanism capable of making a cost-effective contribution to the way we allocate trust is likely to have been harnessed to the task.

The mechanisms involved in epistemic vigilance are of two kinds. Some focus on the source of information and help answer the question: Whom to believe? Other mechanisms focus on content and help answer the question: What to believe?

Whom should we believe, when, and on what topic and issue? We accept different advice from a doctor or a plumber. We believe more easily witnesses who have no personal interests at stake. We are on alert when people hesitate or, on the contrary, insist too much. We take into account people's reputation for honesty and competence. And so on. Much recent experimental work shows, moreover, that children develop, from an early age, the ability to take into account evidence of the competence and benevolence of communicators in deciding whom to trust.[18]

Trust in the source is not everything. Not all contents are equally believable. No one could convince you that $2 + 2 = 10$, that the moon is a piece of cheese, or that you are not yet born. If you were told such things by the person you trust most in the world, rather than taking them literally and at face value, you would wait for an explanation (for example, that $2 + 2 = 10$ in base four, that the moon looks like a piece of cheese, and that you are not yet "born again" in a religious sense). If, on the other hand, the person you trust least in

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

the world told you that a square has more sides than a triangle, that Paris is the capital of France, or that to reach your current age you must have eaten a lot of food, you would agree.

Whatever their source, absurdities are unlikely to be literally believed and truisms unlikely to be disbelieved. Most information, however, falls between these two extremes; it is neither obviously true nor obviously false. What makes most claims more or less believable is how well they fit with what we already believe. If we realize that a claim is incoherent with what we already believe, we are likely to reject it. Still, rejecting a claim that challenges our beliefs may be missing an opportunity to appropriately revise beliefs of ours that may have been wrong from the start or that now need updating.

Vigilance toward the source and vigilance toward the content may point in different directions. If we trust a person who makes a claim that contradicts our beliefs, then some belief revision is unavoidable: If we accept the claim, we must revise the beliefs it contradicts. If we reject the claim, we must revise our belief that the source is trustworthy.

While it is far from failsafe, epistemic vigilance allows us, as audience, to sort communicated information according to its reliability and, on average, to stand to benefit from it. For communicators, on the other hand, the vigilance of their audience diminishes the benefits they might expect from deceiving others—it lowers the chances of success—and if they were caught, it increases the costs they might have to pay in the form of loss of credibility and reputation.

Shouldn't we be vigilant not just toward information communicated by others but also toward the output of our own belief-formation mechanisms? When it comes to beliefs that result from our own cognitive processes, we are the source. Our cognitive mechanisms have evolved to serve us. Unlike other people, these mechanisms do not really have interests of their own. Moreover, there already are, we saw, metacognitive devices that modulate the confidence we have in our own memories, perceptions, and inferences. So, there is little reason to assume that it would be advantageous to apply further vigilance toward the source when we ourselves are the source.

So be it for vigilance toward the source, but what about vigilance toward the content? Wouldn't it be useful to check the coherence of new perceptions and inferences with old beliefs so as better to decide what new beliefs to ac-

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

cept or reject and, possibly, what old beliefs to revise? Actually, coherence checking is not a simple and cheap procedure, nor is it error-proof. Moreover, if it were worth doing, shouldn't its output itself be double- and triple-checked, and so on? If you are willing to distrust yourself, shouldn't you distrust your own distrust? We know of no evidence showing that humans exercise vigilance toward the content of their own perceptions and inferences (except in special cases where they have positive grounds to think that their senses or intuitions aren't functioning well). We know of no good argument, either, to the effect that doing so would be truly beneficial. More generally, people seem to be unconcerned by their own incoherencies unless something—generally someone—comes up to make them salient.

The Argumentative Function of Reason

Precautions have a price. Even when they are on the whole beneficial, they result in missed opportunities. You were wise, for instance, not to pick these mushrooms in the woods since you were not sure they were edible. You were wise even though it so happens that they were delicious chanterelles. Epistemic vigilance is a form of precaution, and it has a price in the same way. A valuable message may be rejected because we do not sufficiently trust the messenger, a clear case of missed opportunity.

Epistemic vigilance, however useful or even necessary it may be, creates a bottleneck in the flow of information. Still, to receivers of information, the benefits of well-calibrated epistemic vigilance are greater than the costs. To communicators, on the other hand, the vigilance of their audience seems just costly. This vigilance stands in the way not only of dishonest communicators but also of honest ones. An honest communicator may be eager to communicate true and relevant information, but she may not have sufficient authority in the eyes of her interlocutor for him to accept it, in which case they both lose—she in influence, he in relevant knowledge.

The more potentially relevant the message you receive from a given source, the more vigilant you should be, but the more costly this vigilance may turn out to be. Here is a dramatic historical illustration. Richard Sorge was a Soviet spy working undercover in Japan during World War II and providing the Russians with much valued intelligence. When, however, he informed them

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

of the imminent invasion of Russia by Nazi Germany in June 1941, Stalin just dismissed his report: "There's this bastard who . . . deigned to report the date of the German attack as 22 June. Are you suggesting I should believe him?"[19] With hindsight, what a mistake! But precisely because the information was both so important if true and so damaging if false but taken to be true, it made sense to be particularly vigilant and indeed skeptical. There are comparable situations in our personal lives where, for instance, we remain skeptical of a warning or a promise precisely because it is too relevant to be accepted on trust.

This is where reasoning has a role to play. The argumentative use of reasons helps genuine information cross the bottleneck that epistemic vigilance creates in the social flow of information. It is beneficial to addressees by allowing them to better evaluate possibly valuable information that they would not accept on trust. It is beneficial to communicators by allowing them to convince a cautious audience.

To understand how argumentation helps overcome the limits of trust, consider what happens when communicators make claims without ever arguing for them. People in a position of power, for instance, typically expect what they say to be accepted just because they say it. They are often wrong in this expectation: they may have the power to force compliance but not to force conviction. Statements of opinion unsupported by any argument are also common in some cultural milieus where arguing, and generally being "too clever," is discouraged. Intelligence, in all the senses of the term, doesn't flourish in such conditions. In both situations, the flow of information is hampered. What argumentation can do, when it is allowed, is ease this flow.

As communicators, we are addressing people who, if they don't just believe us on trust, check the degree to which what we tell them coheres with what they already believe on the issue. Since we all are at times addressees, we are in a position to understand how, when we communicate, our audience evaluates what we tell them. We benefit from this understanding and adjust our messages accordingly. Unless we are trying to deceive our audience, we needn't see their vigilance as just an obstacle to our communicative goals. On the contrary, we may be in a position to use their vigilance in a way that will be beneficial both to them and to us.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

To begin with, even if our audience is reluctant to accept our main point, there may be relevant background information that they will accept from us on trust. By providing such information, we may extend the common ground on the basis of which our less evident claims will be assessed. For instance,

(The doorbell is ringing)
Enrico: It must be Alicia or Sylvain.
Michelle: Actually, Sylvain is out of town. I am sure it is Alicia.
Enrico: Right!

Here Michelle produces a piece of information that Enrico believes on trust: Michelle wouldn't say that Sylvain is out of town if she didn't know it for a fact. This piece of information, however, is a reason for Enrico to revise his conjecture about who might be ringing the doorbell.

A good way to convince addressees is to actively help them check the coherence of your claims with what they already believe (including what they have just unproblematically accepted from you) or, even better if possible, to help them realize that given their beliefs, it would be less coherent for them to reject your claims than to accept them. In other words, as a communicator addressing a vigilant audience, your chances of being believed may be increased by making an honest display of the very coherence your audience will anyhow be checking. A good argument consists precisely in displaying coherence relationships that the audience can evaluate on their own.

As an addressee, when you are provided not just with a claim but also with an argument in its favor, you may (intuitively or reflectively) evaluate the argument, and if you judge it to be good, you may end up accepting both the argument and the claim. Your interlocutor's arguments may be advantageous to you in two ways: by displaying the very coherence that you might have to assess on your own, it makes it easier to evaluate the claim, and if this assessment results in your accepting relevant information, it makes communication more beneficial.

How does a communicator display coherence? She searches among the very beliefs her addressee already holds (or will accept from her on trust) for reasons that support her claim. In simple cases, this may involve a single

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

argumentative step, as in this dialogue between Ben and Krisha, two neighbors of Tang, Julia, and Mary:

Krisha: Tang just called. He is inviting us to come over to their place for dinner with him, Julia, and Mary.
Ben: Mary is there? Then she must have finished the essay she had to write.
Krisha: I would be surprised if she had.
Ben: Surely, if she hadn't finished her essay, she would be working late in the library.
Krisha: But you told me yourself this morning that the library would close early today.
Ben: Ah, yes, I had forgotten. So, you're right—Mary might still not have finished her essay after all but be at home all the same.

Here Krisha uses a piece of information that Ben himself had provided (that the library is closed that evening) as a reason to cast doubt on his claim that Mary must have finished her essay. Krisha's argument shows to Ben that he holds mutually incoherent views, one of which at least he should revise.

Making the addressee see that it would be more coherent for him to agree than to disagree with the communicator's claim may involve several argumentative steps, as in this illustration based on the pigeonhole problem we encountered in Chapter 1 (and which, for once, involves genuine deductive reasoning):

Myra: Since you like puzzles, Boris, here is one. Let me read: "In the village of Denton, there are twenty-two farmers. All of the farmers have at least one cow. None of the farmers have more than seventeen cows. How likely is it that at least two farmers in Denton have the exact same number of cows?" So, Boris, what do you say?
Boris: Well, I say it is likely.
Myra: I say it is certain!

Myra and Boris have come to different conclusions, and neither has the authority to persuade the other just by saying, "It is so." Still, where authority fails, argument may succeed.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Myra: Well, imagine you go to the village, gather the twenty-two farmers, and ask them to stand in groups, each group made of farmers who have exactly the same number of cows.

Boris: How does that help? It could be that there are no farmers who have exactly the same number of cows, and then, in each of your "groups," there would be only one farmer.

Boris's reply has an entailment that he is not yet aware of but that Myra highlights:

Myra: This is impossible. Since none of the farmers have more than seventeen cows, there couldn't be more than seventeen groups: a group for farmers with one cow, a group for farmers with two cows, and so on up to a group for farmers with seventeen cows. But then, given that there are twenty-two farmers and at most seventeen groups, there would be at least one group with several farmers, right?

Boris: Yes, and?

Myra: And since farmers in the same group would have the same number of cows, then it must be the case that at least two farmers have exactly the same number of cows. This is certain, as I said, not merely likely.

Boris: You are right. I see it now.

What Myra does is present Boris with intuitive arguments that spell out some clear implications of the situation described in the puzzle. The conclusion of the last intuitive argument, however, directly contradict Boris's initial conclusion. When he realizes this, it is more coherent for him to agree with Myra and change his mind.

The dialogue between Myra and Boris is a trivialized version of the Socratic method: help your interlocutors see in their own beliefs reasons to change their views. When reflection on reasoning began in the Western tradition in the work of Plato and Aristotle, the argumentative use of reasons to try to convince an interlocutor, a court, or an assembly was seen as quite central. The social aspect of reasoning was well in evidence. Socratic reasoning could be seen as reasoning par excellence. Then, however, starting already in the work of Aristotle, the study of reasoning took a different turn.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

Important normative questions about what makes an argument valid led to a more abstract approach to reasoning and to the development of logic proper. The use of reasons in argumentation could now be seen as just one practical application among others of more general reasoning principles. A new image of the typical reasoner emerged. Rather than Socrates trying to convince his interlocutor and the interlocutor understanding the force of Socrates's argument, the paradigm of a reasoner became the scientist reasoning on his own (more rarely on *her* own) to arrive at a better understanding of the world. From the point of view of the psychology of reasoning, this has been unfortunate: it has obscured the degree to which reasoning (including scientific reasoning) is a social activity, and the degree to which it is based on intuitions.[20]

Reasons are not arbitrary rhetorical devices. If they were, how would they have any force? Reasons are supported by intuitions that are themselves based on genuine cognitive competencies. Intuitively good reasons are more likely to support true conclusions. Imagine for instance that instead of Myra trying to convince Boris, Boris would have tried to convince Myra that his answer was the right one. How likely would he have been to succeed? Not very. His reasons would have been intuitively wanting. In fact, in trying to formulate them, he might himself have become aware of their inadequacy. In this case, what Myra and Boris are disagreeing about is a logical problem with a demonstrable solution. Even when the question on which two people disagree does not have a demonstrably true answer, there may be intuitively much better argument for one answer than for another. Then, of course, there are cases where two mutually incompatible conclusions can each be supported by plausible but inconclusive reasons. In such cases, argumentation by itself typically fails to change people's minds.

We construct arguments when we are trying to convince others or, proactively, when we think we might have to. We evaluate the arguments given by others as a means—imperfect but uniquely useful all the same—of recognizing good ideas and rejecting bad ones. Being sometimes communicators, sometimes audience, we benefit both from producing arguments to present to others and from evaluating the arguments others present to us. Reasoning involves two capacities, that of producing arguments and that of evaluating

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

them. These two capacities are mutually adapted and must have evolved together. Jointly they constitute, we claim, one of the two main functions of reason and the main function of reasoning: the argumentative function.[21]

Main Functions, Secondary Functions, and Sea Turtles

The central thesis of the book is that human reason has two main functions corresponding to two main challenges of human interaction: the attribution of reasons serves primarily a justificatory function, and reasoning serves primarily an argumentative function.

Why do we say the two *main* functions rather than just the two functions? Frankly, out of prudence. We are not convinced that the reason module has any other function than these two, but we want to leave the possibility open.

We have not denied, for instance, that attributing reasons to an individual may be done for explanatory rather than justificatory purpose. One may attribute reasons without assessing their justificatory value. This is, after all, what historians do when they seek to explain the actions of people from the past in a nonjudgmental way. Couldn't such explanatory use be beneficial? Couldn't it be at least a secondary function of the attribution of reasons?

We have not denied that reasoning can be pursued individually, either to produce arguments aimed at convincing others or to evaluate others' arguments. Couldn't such individual inquisitive reasoning be beneficial? Couldn't producing such benefits be, if not the main function, at least a secondary function of reasoning?

Well, possibly, but to better understand what it would take to establish such claims, consider sea turtles.

Although sea turtles are descendants of land reptiles, they spend all their time in the water, or almost all their time. Females lay eggs out of the water, in nests they dig on the beach, where therefore all sea turtles are born (and then try to get into the water as fast as possible).

The limbs of sea turtles have become exquisitely adapted to life in the sea: their forelimbs have evolved into flippers and their hind limbs into paddles or rudders. What could be better for swimming, which is the ordinary mode of locomotion of sea turtles? On repeated occasions, however, females must

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

use these same limbs to crawl out of the water and to dig nests in the sand. The use of limbs in such a manner, even if rare and clumsy, plays a direct role in reproduction and is clearly adaptive. To argue conclusively, however, that digging nests is a secondary function of sea turtle limbs, one would have to find features of these limbs that are best explained by this use. Otherwise, this might be best described as a beneficial side effect.[22]

Female turtles' nest digging is obviously beneficial, even if the limbs they use are not specifically adapted to the task and if their performance is notably clumsy. Similarly, if reason has beneficial effects different from its two main functions, these might correspond to secondary functions. If so, we would expect to find features of reason tailored to the achievement of these benefits. Otherwise, we might suspect that these are mere side effects of reason.

It is plausible that the capacity to explain people's ideas and actions by attributing to them reasons is on the whole beneficial (even if, as we saw, it is at best a distortion of the psychological processes involved). It is less obvious but not inconceivable that genuine individual inquisitive reasoning (as opposed to mentally simulated argumentation) does more good than harm in guiding beliefs and decisions. Still, even assuming that these two uses of reason are each, on average, beneficial, it does not follow that they are, properly speaking, functions of reason. To argue that they are, one would have to find some specific features of the attribution of reasons on the one hand and of reasoning on the other hand that are geared to the production of these particular benefits. We are not aware of any such evidence. In the case of reasoning, what is more, it is not just that features specifically tailored to the fulfillment of its alleged inquisitive function are lacking; it is that several well-evidenced features of reasoning would actually undermine this function.

We should keep an open mind regarding possible secondary functions of reason, of course, but at present, the challenge is to establish what beneficial effects would explain why reason evolved at all—in other terms, it is to identify reason's primary function or functions. In trying to answer this challenge, we denied that anything like classical Reason, with the capacity to procure better knowledge and decisions in all domains, has ever evolved. What has evolved rather is a more modest reason module—one intuitive inference

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

module among many—specialized in producing intuitions about reasons in the service of two functions, justificatory and argumentative.

In Parts IV and V of this book, we will demonstrate, with evidence rich and varied, that reason is precisely adapted for fulfilling these two functions. We will show, moreover, how this new interactionist approach illuminates the role reason plays in human affairs.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.

EBSCOhost: eBook Collection (EBSCOhost) printed on 8/8/2026 7:17:40 PM UTC via TULANE UNIVERSITY. All use subject to https://www.ebsco.com/terms-of-use.