Review of Mercier and Sperber's *The Enigma of Reason*

By Jeffrey Maynes, St. Lawrence University

Abstract

In *The Enigma of Reason*, Hugo Mercier and Dan Sperber (2017) defend the proposal that reason is a specialized module which produces intuitions about reasons. Reason serves two functions: for individuals to justify their own judgments and actions to themselves and others, and to persuade others. In this review, I briefly summarize the central claims of the book, critically examine Mercier and Sperber's arguments that reason is not a general faculty underlying our inferential abilities, and explore the pedagogical implications of their work.

Keywords: reasoning, pedagogy, inference, argument

Introduction

In *The Enigma of Reason,* Hugo Mercier and Dan Sperber (2017) attempt to develop and motivate a novel account of reason, according to which reason is not a faculty that we employ in domain-general problem solving, nor is it the underlying structure of thought. Rather, reason is an intuition-producing system, one that produces intuitions about reasons. In particular, it provides a *post hoc* rationalization of our judgments. The purpose of such a faculty is fundamentally social; it has evolved to help us justify our judgments to ourselves and persuade others of the rightness of our judgments. In order to understand how reason works, and its virtues, they argue that we need to look at human beings in a social context.

In this review, I first provide a brief overview of the central claims of the book. Throughout, Mercier and Sperber develop an inference to the best explanation, arguing that their account of the nature and function of reason best explains a wide array of results from psychology and cognitive science. I then turn to two sets of critical remarks. First, I raise concerns about Mercier and Sperber's argument against understanding inference in terms of unconscious logical reasoning. Second, I raise a series of questions inspired by their work for future research in reasoning pedagogy.

Summary

The titular enigma arises from a tension in our conception of reasoning as a flawed superpower. On one hand, it is seen as what distinguishes humans from other animals and, when implemented properly (or when truly implemented without interference), leads to logically rigorous, truth-preserving inference. Viewed through an evolutionary lens, they contend, such a superpower is mysterious. How could it evolve, and why is it so rare? What intermediate steps, themselves adaptive, could have led to its development?

On the other hand, this superpower seems to be highly flawed. Psychologists and others have assembled an impressive array of evidence that we are biased reasoners, prone to making regular and frequent errors. Since such errors would seem to be maladaptive, why would reason have evolved to regularly make such errors? The difficulty in need of explanation is: if reason is a superpower, how did it evolve? If it is flawed in these ways, why did it evolve as it did? This approach drives the book. By putting reason in an evolutionary framework and understanding its function, they argue, we can better understand what reason

is, and why it is so powerful. Its function is to recognize and identify reasons for claims, so that we can use these reasons to advance arguments with the purpose of convincing others, and to justify our own beliefs and decisions to ourselves and others. This view, which Mercier and Sperber call *interactionism*, is contrasted with *intellectualism*. For the intellectualist the function of reason is to allow individuals to draw truth-preserving inferences, and so to form true beliefs about the world.

The book is divided into five sections. The first sets out the enigma, and situates Merceir and Sperber's evolutionary approach within the field. Section two gives an account of inferential modules, and lays the groundwork for thinking of reason as such a module. The third section develops this idea, arguing that reason is a form of intuitive inference about reasons for claims. Section four then develops the social function that such a module plays, bolstering both the evolutionary and epistemic case for reason. The final section applies this account to reasoning 'in the wild,' such as in the cases of political and scientific reasoning. Here I will focus on the position Merceir and Sperber put forward in sections two through four, beginning with their account of inference.

Mercier and Sperber define inference as "the extracting of new information from information already available" (p. 53). This characterization is broader than that found in logic textbooks, where inference is typically defined in terms of the logical processes or steps involved in extracting that information. That is, on the traditional view, if I conclude that 'my car needs repairs' from 'if my check engine light is on, then my car needs repairs' and 'my check engine light is on,' it is the fact that I have used *modus ponens* that makes it an inference. On Mercier and Sperber's view, however, the nature of the intermediate steps are irrelevant to it being an inference. Indeed, this conception of inference is so broad that they classify perspiration as an inference (taking information about body temperature as input, and instructions to the sweat glands as outputs).

These inferences are often performed by specialized mechanisms. Though Mercier and Sperber are skeptical of the significance placed on the Wason Selection Task in the psychology of reasoning, it provides an apt illustration. When the task is presented in abstract logical form, most respondents, no matter their education level, fail. Yet, as Cosmides famously found, when the case is presented as a norm violation (with identical logical form), respondents tend to get it correct (Cosmides, 1989). If we applied the same inferential mechanism to both cases, then we would expect that we would reach the same answer in both cases (if not the correct one). Instead, perhaps, we have an inferential mechanism that specializes in norm violations, one that is not invoked by the traditional problem.

Such inferential mechanisms can range from those involved in facial recognition, to the learned mechanisms involved in reading. What unites them is that they involve inference (such as extracting the meanings of words from the visual data in reading), that they are task-specific, that they are fast, and that they are automatic. These mechanisms are modular in the sense that they are "autonomous mechanisms with a history, a function, and procedures appropriate to its function" (p. 73). Much like Mercier and Sperber's account of inference, this is a broad definition. Indeed, the entire brain counts as a module on their approach. This broad definition is intended to avoid some of the commitments of a Fodorian conception of modules, such as informational encapsulation (Fodor, 1983).

Inference does not require reasons in the sense of *psychological reasons,* or a representation of a fact that supports some conclusion, as a reason. This final condition is crucial. A representation that causes me to infer some piece of information is not a reason unless I represent it as a reason, because otherwise we could not explain why we draw the conclusions we do from the representation of a reason, since that fact is also a reason for

an indefinite number of other conclusions. This is a demanding representational requirement; having reasons for your beliefs requires metarepresentational capacity. I discuss this argument in more detail below.

Of course, we are able to provide reasons for our judgments when asked. Might it be that the fact that we can report these reasons is evidence that those reasons are, in fact, the reasons we used to reach the conclusions we did? Mercier and Sperber, however, point to the evidence that we often do not know the causes of our own behavior and judgment (most famously, Nisbett & Wilson, 1977). The reasons we give are often *post hoc* rationalizations, not recall of an implicit, actual reason for our judgment. These rationalizations are the product of the reasoning module. Reasoning is a "use of intuitive inferences about reasons" (p. 133). Like other inferential mechanisms, reasoning is a specialized module. The output produced by the reasoning module is of the form: 'R is a reason for C.' It is intuitive in that we do not know why we judge that R is a reason for C. It is an inference in that it is the extraction of new information (that R stands in the reason relation to C) from other information (R and C).

Why think that reason is a form of intuitive inference? As noted above, for something to be a psychological reason, we need to represent it as a reason. An infinite regress looms if I need a further reason to recognize R as a reason for C. After all, then I will need the further metarepresentation: 'R1 is a reason to recognize R2 as a reason for C.' What, then, is my reason to see R1 as a reason? While such cases are possible (scholars often care about the reasons for our reasons), to avoid the regress, our inferences about reasons need to bottom out in inferences operating without appeal to reasons.

In some ways, reason is like the press secretary for our inferential systems. Suppose that a politician takes a position on immigration, and asks his/her press secretary to defend it. The press secretary may not know why the politician took that position, but will do his/her best to find justifying reasons for it that will enhance the politician's reputation and convince others to adopt that position as well. If, for example, the politician was driven by racial animus, and the press secretary knew that such a reason would harm the politician's reputation, s/he may articulate a more socially acceptable reason. Similarly, the reason is not involved in the decision making, and may not have any access to the actual causes of our judgments. Its role is to take those judgments and identify reasons which would justify them for public consumption.

This is not, however, to say that there is no role for reflective reasoning. The reasoning module produces an intuition about the reason relation holding between R and C. It indirectly, however, gives us reason to believe C. If I draw conclusion C on the basis of my intuitive inference that R is a reason for C, I am engaged in reflective reasoning since I am aware of the reasons for believing C. Reflective reason turns out to be an indirect output of the intuitive inferences produced by the reason module, not a separate faculty.

For example, suppose that we are discussing who to vote for in an upcoming election. I support candidate A, and you support candidate B. I believe strongly that health care is a right (so strongly, in fact, that this issue typically determines how I vote), and you point out that candidate B's plan is more likely to promote access to health care among the citizenry. My reasoning module recognizes this as a reason to vote for candidate B. Upon so recognizing it as a reason, I conclude that, despite my initial views, I should vote for B. Here I have reflectively reached the conclusion that I should vote for B, because I have considered the reasons for doing so, based on my intuitive inference that this was, in fact, a reason to vote for B.

Why would such a reasoning module have evolved? It has two main functions. First, the attribution of reasons plays a justificatory role that matters to building and maintaining

a reputation. Second, giving reasons plays an argumentative role in changing the minds of others. Unlike the intellectualist approach, these goals are essentially social; reason's function is not to help individual reasoners to identify truths reliably, even if, on occasion, it helps them to do so. Here I will focus on the argumentative role, and Mercier and Sperber's argument that the interactionist approach better explains myside bias (the tendency to seek out and favor evidence supporting your position). If the goal of reasoning is for individuals to find the truth, then why did it evolve to systematically lead us away from truths? After all, this bias seems to incline us towards defending our prior preconceptions, even when they are wrong. If the function of reasoning is to promote individual truth-seeking, we should have a bias towards falsifying information. Put into the interactionist picture, however, Mercier and Sperber argue that the bias makes sense. Indeed, it is a useful adaptation of our reasoning system. We tend to seek out evidence for our own positions because our goal is not to find truth, but to persuade others, and developing our case as strongly as possible is more convincing.

The goal of changing minds is not only advantageous for pragmatic reasons but, in the right social context, is epistemically advantageous as well. While I myself may not be able to make the best case for opposing positions, other people in my community will. As a community we will better approximate the truth, as individuals committed to defend their positions will give us a greater variety of well-defended positions to consider. Myside bias works on analogy with an adversarial legal system. A single judge, tasked with considering all possible positions, may fail to fully develop each position. If the judge is subject to any cognitive biases inclining her in one direction, as we all are, this risk is even more pronounced. If, on the other hand, lawyers are tasked with making the case for each side as best they can, their vested interest in defending that position will lead them to find the strongest arguments for their side that they can muster. The court is in a superior epistemic position because the various sides of the issue will be defended with the better arguments, even though each individual lawyer may be providing a one-sided account.

We are, they contend, lazy reasoners, in the sense that myside bias inclines me to make a case for my position P, and I am inclined to stop defending it once I have a reason for it. Others, however, are epistemically vigilant, and are good at evaluating the arguments of others. If you remain unconvinced by my reasons, then I will have to come up with more reasons to change your mind (and preserve my reputation). This laziness is a "sensible" laziness (p. 236). When coupled with the epistemic vigilance of others, we devote resources to defending positions only when we need to in order to justify ourselves or convince others. It is cost-efficient to be individually lazy, and collectively vigilant. A central prediction of this view (for which Mercier and Sperber offer preliminary supporting evidence) is that we will be better at evaluating arguments than we are at producing them.

The motivating enigma of this book is how to explain the evolution of a flawed reasoning ability. If its function is to help us discover truths about the world, why did it evolve to so systematically mislead us? This enigma dissolves when we recognize the real function of reasoning - to convince others and to justify our positions. Indeed, reasoning is highly adapted to this function. Biases like myside bias help us achieve it, and what's more, when situated in the right context, reason functions *better* than in the individualistic context.

Critical Reflections

Before turning to my critical remarks, it is worth being clear: this is a book that anyone interested in the nature and psychology of human reasoning should read. It offers a novel

and thought-provoking analysis of the function of reason, and raises important questions about the role of evolutionary psychology in understanding it. Further, it is an engaging read, one that would be accessible to a wide audience, including scholars across the various disciplines interested in its central questions.

First, a general note. Mercier and Sperber cover an enormous amount of ground in this book. As such, there are a number of points where the psychological evidence merits extended analysis and discussion. Indeed, many of the studies discussed have themselves been the subject of a great deal of discussion in the literature. Given the scope of Mercier and Sperber's theory, this is not surprising. Many of the chapters of this book could themselves be books (and many of the topics are indeed the subjects of several books), and Mercier and Sperber's goal is a synthetic approach that makes sense of ongoing psychological research in a number of related domains. As such, this book is best understood as motivating a hypothesis, rather than providing a definitive case for it. In my critical remarks below, I indicate just some of the areas of the book that merit further work, with a focus on the role of logical principles in how we draw inferences, and the pedagogical implications of the arguments of the book.

Reasoning and Inference

As noted above, one alternative interpretation of the nature of reasoning is that it is a faculty that underlies a wide range of inferential capacities. For example, if I employ a module devoted to drawing inferences about norm violations, that inferential module might still make use of *modus ponens* reasoning to produce its output. In such a case, the reason relationship is not a *post hoc* determination by a reasoning faculty, but rather a description of the processes actually employed in the original inference. This is not to deny that there are *post hoc* rationalizations as well, or to say that we have introspective access to all of our reasons. It is, rather, to suggest that reasoning may indeed be a domain-general faculty that underlies even highly specialized inferences.

Mercier and Sperber argue that representations cannot be reasons unless they are a particular kind of metarepresentation: a representation of another representation as a reason. This is required to solve the problem of explaining why we draw the conclusions we do from reasons. After all, an objective reason, a fact supporting some conclusion, will support multiple conclusions, sometimes even contrary ones. For example, "the fact that it has been snowing may be a reason to stay at home and also a reason to go skiing" (p. 119). Simply pointing to this fact, or a representation of it, does not explain why one individual goes skiing and another remains at home. The difference must be that one individual represents the fact as a reason to go out (and *mutatis mutandis* for going skiing). This objection is a version of the frame problem (Shanahan, 2016). If thought is a logically structured language, why do we not simply draw out every logical consequence of our beliefs?

Mercier and Sperber argue that most of our inferences do not involve these kinds of metaprereresentations. For example, desert ants, after taking a circuitous path in search of food, are able to then move in a straight line back to their nest. They do so, not on the basis of complex mathematical reasoning, but by a process of integrating information about distance traversed with information about the polarization of the sun's light. Given the task-specificity of this ability, and what we know about ant brains, it is more likely that the ants are using a specialized inferential mechanism that is not rooted in a general logical competence that could be applied to other problems.

Similarly, the reactions of four-and-a-half-month-old infants to physically impossible events (such as a box floating in the air after its supports have been removed) suggest that they were surprised by those events. One could

explain their surprise by appeal to a logical inference concluding that the box will fall. Just as with the ants, however, Mercier and Sperber contend that it is implausible that these infants are employing the slow logical reasoning of relying on syllogisms.

Mercier and Sperber then generalize, arguing that it is unnecessary to postulate that inferences happen through logical reasoning for most of our inferences. Instead, we can explain inferences by appeal to procedures that apply to particular kinds of representations. These procedures can be highly task-specific, and exploit regularities in their environment. For example, a module devoted to recognizing snakes may assume that objects with certain snake-like properties (e.g., shape, position, movement) are, in fact, snakes. I can infer directly from the snake-like properties to 'there is a snake here' without appealing to general premises about snakes. That is, we can explain my fearful reaction to a snake (or coiled garden hose) as the operation of an associational procedure that produces that fearful reaction based on the representation (of something having snake-like properties) it took as input.

This approach has several advantages, they argue, over the logical reasoning view. In particular, the logical reasoning view leaves open four important questions that Mercier and Sperber's model shows better likelihood of solving. First, how did this general logical reasoning faculty evolve? Second, how does it develop in individuals? Third, why does it only provide relevant inferences, instead of drawing out every logical consequence of a set of premises? Fourth, why do people come to divergent conclusions from the same information, if relying on the same logical principles? Since, Mercier and Sperber suggest, each of these questions is better answered by their view, we need not postulate the existence of an underlying logic system. One reason one might think that these inference modules are actually drawing logical inferences is that we need to explain how the new information is produced. While in some cases associational mechanisms might be sufficient, the complexity of other cases seems to suggest that we are doing things such as employing *modus ponens* or disjunctive syllogism. Mercier and Sperber's reply shares much in common with Michael Devitt's argument against taking linguistic rules to be either represented or embodied in minds or brains (Devitt, 2006). Devitt argues that linguistic rules provide constraints that any account of how language works ought to respect, rather than actually being the rules used in the cognitive processing behind language use. Similarly, logical rules may be constraints for inferential modules to respect in certain conditions without their being implicated in the inferences themselves. That is, our inferential modules produce outputs that *look* like *modus ponens,* because both the module and *modus ponens* are reliable ways to generate information, but do not actually employ *modus ponens.*

Mercier and Sperber's case would be well served, however, by attending more to research in animal psychology. While far from conclusive (as I will note below), whether animals perform logical inferences has been the subject of a number of important studies. Inference by disjunctive syllogism in non-human animals is most famously studied through the two cups task (Call, 2004). In this task, a subject is shown two empty cups. A piece of food is placed into one of the cups, and while the subject knows that food has been placed in a cup, they do not know which cup now contains food. The subject is then given information that reveals which cup is empty, before being allowed to choose between the two cups. If the subjects choose the cup that was not revealed to be empty, this suggests that the subject is employing disjunctive syllogism to identify the location of the food. Apes, siamangs, olive baboons, capuchin monkeys, lemurs, dogs, ravens, carrion crows, and African grey parrots have all been shown to complete this task successfully (Mody & Carey, 2016).

Alternative interpretations of the significance of the two cup experiment are available (see Beck, 2018). One variant, designed to better test whether subjects are using disjunctive syllogism, is the four cups task (this approach is designed to rule out the possibility that subjects use an 'avoid the empty' heuristic). In this task, the subject is presented with two sets of two cups (A/B and C/D), and a reward is placed in one cup in each set. The subject can see that a reward has been put into one of these two cups in each set, but is not aware of which cup. An experimenter then reveals that one of the four cups is empty, and subjects are presented with a choice of cups. Suppose that the experimenter has revealed that cup A is empty. If the subject uses disjunctive syllogism, she should chose cup B (the other cup in the set with A). If she does not, then she may also pick cups C or D. Mody and Carey ran this test with children, ages 2.5 years and 3-5 years (Mody & Carey, 2016). They found that the choices of the 2.5-year-old children were not consistent with use of disjunctive syllogism, but the choices of children aged 3-5 were. This provides some reason to believe that these kids are, in fact, using disjunctive syllogism to solve this task, contra Mercier and Sperber.

Indeed, that there is evidence of the use of disjunctive syllogism in non-human animals, and of how it develops in children, suggests that the logical reasoning approach could plausibly answer the first two of Mercier and Sperber's four questions. The presence of the ability in some animals might provide a window into how and why it evolved in humans, and its development in children is a first step towards an account of how it develops in individuals. Further, while task-specific modules are evolutionarily parsimonious with regard to success on the task in question, a domain-general logical faculty may be more parsimonious than a full accounting of all of the task-specific modules required to navigate the world. Whether this approach will answer Mercier and Sperber's four questions better than their own account remains to be seen, but it is far from obvious that these questions cannot be satisfactorily answered on the traditional view. Perhaps, as with Devitt's arguments about linguistic knowledge, this is not evidence that disjunctive syllogism is actually being used. Rather, it is evidence that the subjects' inferences produce the same behavior as inferences by disjunctive syllogism. That we can *describe* the activity in logical terms does not imply that the logical structure is implicated in the mechanism. In defense of attributing logical structure to non-human animals, Tyler Burge writes:

> The propositional structure is very simple, and the exclusion transitions themselves seem to match that structure, element for element. The sequence of entertaining the alternatives, rejecting one, and choosing the other mirrors and matches the structure of the propositional inference by exclusion. Together with the first two considerations, it seems to me that this fact could favor - as a best explanation - the attribution of logical structure to the psychological states of non-linguistic animals. (Burge, 2010, p. 65)

That is, because the process of entertaining options, eliminating one, and selecting the remainder matches the logical structure of disjunctive syllogism, we have reason to believe that the psychological states are logically structured. Unlike structurally different but mathematically equivalent descriptions of some behavior, Burge argues, the disjunctive syllogism cases show a matching structure between the inferential process and the logical rule. If so, then this matching serves as a bridge premise, allowing us to infer from the logical rule as a description of behavior, to the rule as part of the cognitive processing behind the behavior.

Another advantage of considering (purported) logical inference in non-human animals is that such animals plausibly do not have metarepresentational states (Proust, 2018). If so, then a consequence of Mercier

and Sperber's view is that non-human animals do not have reasons for their actions. They may have *causes* for their actions, and evolutionarily adapted mechanisms for inferring appropriate courses of action from other information, but the causes or mechanisms do not count as reasons. Yet, if nonhuman animals do make use of a general logical principle such as disjunctive syllogism, then plausibly they do have reasons for their actions even without metarepresentational capacity.

This is ultimately a complex open question, and my aim here is not to provide a decisive refutation of Mercier and Sperber's approach. Rather, it is to suggest fruitful directions for further work, and to note complexities which their account will need to address. The modular inference approach, with a series of domain-specific heuristics and particular associations, may be capable of explaining a great many judgments we make. The challenge facing that approach is to show that this approach is sufficiently powerful that it is likely that all of our inferences work this way, and to give an account of these inferential procedures that does not appeal to logical rules like disjunctive syllogism.

Pedagogy

What, if anything, should critical thinking educators take out of this work? First, a word of caution. As noted above, Mercier and Sperber's arguments are best interpreted as motivating a hypothesis. Further work is required before educators should adopt their model of reasoning as the basis for pedagogy. That said, debates over the nature of reasoning highlight the complex interaction between pedagogy and theory. If our pedagogy is grounded in false or incomplete theory, we risk it being ineffective, or worse, harmful.

One plausible response is to retreat to informal and formal logic. These fields focus on the contents of arguments, and are neutral with regard to the underlying psychology. Our psychology may make us more or less likely to endorse cogent or fallacious inference under certain conditions, but that psychology is not what makes the arguments cogent or not. The flaw with this approach, as I have argued elsewhere (Maynes, 2013, 2015), is that it is unlikely to promote the goal of teaching students to be better critical reasoners, and not merely to gain a better understanding of the content of informal and formal logic. The best path forward is to identify effective pedagogical interventions that are grounded in well supported components of our psychological understanding, and otherwise suggest directions for future research in reasoning pedagogy.

Perhaps the most important claim Mercier and Sperber make for critical reasoning pedagogy is that reason is a module producing intuitions about reasons. Anecdotally, this explains a persistent struggle with teaching reasoning: reconstructing arguments requires a recognition of what information is evidentially relevant, and what is illustrative, extraneous, etc. This is particularly a challenge with rhetorical devices that have persuasive effect without having evidential force. We would expect this to be the case on Mercier and Sperber's approach, as the reasoning module evolved in a social context. Reasons are successful when persuasive, and so we would expect to be better at recognizing reasons-as-persuaders than recognizing reasons-as-evidence. Further, if recognizing reasons is intuitive, it will be difficult for educators to fully articulate why they identify something as a reason in the first place.

Developing the ability to recognize reasons precedes the ability to logically construct inferences in pedagogy, *even if* the logical inferences are why the reasons are, in fact, reasons. This is because students first have to constrain the problem space to a set of plausible solutions. One cannot simply test every possible arrangement of premise/conclusion relationships for all of the sentences of a text (let alone implicit

premises). Instead, students need to be able to recognize reasons and conclusions, at least in a rough form, and to then apply their knowledge of argument forms to articulate and make precise the structure of the argument. The first task for further research is to determine how our intuitions about reasons can be improved. Are there heuristics that students might follow (such as when critical thinking textbooks provide students with keywords) until our intuitions improve? Or, is it a matter of apprenticeship, where students work with arguments alongside a skilled reasoner (the instructor), and come to improve their intuitions through observation and imitation? Mercier and Sperber suggest having students exchanging arguments, helping them to anticipate counter-arguments, and so develop a better sense of what reasons will work. Even then, care is needed to ensure that students' reasoning abilities do not develop to identify *persuasive* claims as reasons, rather than *evidential* ones (lest they become *mere* rhetoricians).

Mercier and Sperber's account does, however, give some hope that developing these intuitions provides an answer to the challenge of domain generality (e.g., Willingham, 2007). The reason module provides a form of "virtual" domain generality through reflective judgment (p. 182). Recall that reflective judgments are reached when we infer from the intuition 'R is a reason for C' to the truth of C. Improved reasoning skill may not improve the intuitive inferences made by other modules, but by improving our recognition of reasons, and so making us more likely to actually draw conclusions based on that recognition, we may inculcate better reasoners across a wide range of domains by teaching critical reasoning skill.

Mercier and Sperber's account also raises crucial questions about our ambitions to debias our students. One common goal of critical reasoning courses (and one I have explicitly argued for, see Maynes, 2013) is to help students mitigate the effects of cognitive biases such as myside bias. Mercier and Sperber, along with others working on ecological rationality (Gigerenzer, 2008), develop a competing picture, where debiasing may in fact be harmful (if not utterly ineffective). These biases, they argue, are actually rational, when understood in the proper context. Myside bias may lead us into error when reasoning as lone individuals, but may lead to better outcomes when used in a social context where others will champion alternative views. Just as teaching baseball players to catch fly balls using calculus will lead to be less effective outfielders, so too debiasing students may lead them actually to be *less* effective reasoners! Consider, for example, the recognition heuristic. Following this heuristic, one would use the recognizability of the options as the criteria for making a choice between them (Gigerenzer, 2008). In one famous study, German participants more accurately judged the relative population size of US cities than did American participants, likely because they ranked the cities based on how recognizable they were to them (Goldstein & Gigerenzer, 2002). If students were sufficiently debiased, such that they did not make use of the recognition heuristic, and instead engaged in slow, reflective deliberation, they, like the American participants, may make *worse* choices in some cases. They would be prone to making worse choices because the heuristic, Gigerenzer argues, is rational in some circumstances, and has a decisive advantage in speed and cognitive demand.

Here I will suggest two directions we might go in response to this challenge. One is to abandon the goal of mitigating bias in the production of arguments. This, however, might be coupled with strategies for articulating opposition positions and anticipating them in our own arguments, as well as instruction on the value of actually testing arguments by pitting them against opposing arguments. We may anticipate these opposing arguments for the purpose of refuting them (myside bias in action), but this approach might still encourage us to think of the reasoning process as

essentially collaborative, rather than essentially individualistic. Further, if, as Mercier and Sperber predict, we are better at evaluating arguments than producing them, then teaching effective strategies for recognizing how bias will lead to problems in other arguments will help students develop a skill that is more amenable to improvement.

In fact, put into practice, this approach likely has a lot in common with how critical thinking is currently taught. Teaching students fallacies and tools for argument analysis helps with the evaluation of arguments, and actually engaging peers in substantive debate exercises helps them better anticipate, and respond to, objections. We may be conceiving of this pedagogy wrongly in intellectualist terms, but it still may be effective in producing the kinds of reasoners who will engage in social deliberation effectively.

The second approach, which I defend (Maynes, 2017), instead sees the aims of critical-thinking education in terms of developing the metacognitive skill to recognize when certain kinds of debiasing strategies are effective, and use them in those contexts. That is, on the ecological rationality approaches to the cognitive biases, there are certain conditions wherein our biases tend to be rational (such as reasoning in dialogue with others, being in cases where decisions are needed quickly, or being in cases where certain information is limited). Yet, these conditions do not always obtain. We are often presented with high stakes reasoning scenarios where these biases do indeed lead us astray. The contemporary political environment provides an apt example, where we often reason in social contexts where our interlocutors hold the same positions we do. The myside bias may lead such insular epistemic communities to reinforce the popular view (since few people will defend the alternatives), even if it is effective in other contexts. The skilled critical thinker is someone with the metacognitive knowledge and skill to recognize the relevant features of their environment, and employ the appropriate cognitive strategies for that context. On this approach, we ought to be attempting to mitigate biases like myside bias by helping students not only develop the cognitive strategies to reduce those biases, but to recognize when to employ them.

Mercier and Sperber also raise worries about teaching fallacies, and though sympathetic with their aims, I will argue that their case is unsuccessful. They write that the problem with fallacies is that "for almost each and every type of fallacy in such lists, there are counter examples in the form of arguments that meet the criteria to be declared fallacious but that in real life are quite acceptable or even good arguments, arguments that might convince a rational audience" (p. 229). They offer two such counter-examples:

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

Ad Ignoratiam
Policeman: I am convinced that Ishii is a law-abiding citizen. I could find no evidence that he is not.

First, it is worth noting that neither of these are counter-examples. In the first case, Makiko's reply is effective insofar as it challenges the credibility of Yoshi's testimony. Yoshi does not offer any explicit evidence for the claim that the chocolates aren't very good, and so is plausibly interpreted as appealing to his own experience in tasting them. That is, the reason to believe that the chocolates are not very good *is that Yoshi said so.* In that case, it is not an instance of *tu quoque* to challenge the credibility of that testimony. If, however, Yoshi said that they are not very good, and pointed to the reputation of the company that made them or the advice of other chocolate experts, then the fact that he ate almost the entire box is irrelevant (perhaps he was too weak to avoid

even bad chocolate), and Makiko committed a fallacy. In the *ad ignoratiam* case, Mercier and Sperber contrast the example above with one where the policeman concludes that Ishii is *not* a law-abiding citizen, because no evidence could be found that he is one. Both arguments are equally fallacious. The argument that he is a law-abiding citizen only appears successful because it accords with our commitment to the presumption of innocence. We do not actually have evidence that Ishii is law-abiding; it is simply that we have a moral norm in favor of assuming that he is when there is no evidence to the contrary.

Second, that these arguments are deemed acceptable in real life does not indicate that they are not fallacious. Plenty of arguments are widely accepted despite their fallacious nature. Indeed, this is why people teach the fallacies! Even if one grants Mercier and Sperber the idea that the central aims of argument are justificatory and persuasive in a social context, the fallacies are descriptions of logical problems with the cogency of arguments.

That said, I remain sympathetic with their aims here. Fallacies are slippery outside of stock examples, as whether an argument is fallacious depends upon how it is reconstructed, and there are often charitable readings of arguments that avoid the fallacy. The key question for future work is not whether these cases are counter-examples to the fallacies as such, but whether teaching the fallacies encourages students to recognize fallacies only when they occur, or if it reinforces students' tendency to overgeneralize in ways that make them less effective reasoners.

Finally, Mercier and Sperber's approach merits further work on the nature and success of teaching social reasoning strategies. Traditionally, critical thinking courses are taught to help individual students reason better, and are assessed with this goal in mind. According to Mercier and Sperber, however, not only have we evolved to reason in social scenarios, we are more *epistemically effective* when reasoning in groups. If our goal is to help students develop truth-maximizing skills, then developing the skills to reason effectively in conversation with peers may best achieve that goal. Indeed, regardless of whether Mercier and Sperber are right, this is a worthy opportunity for further work. Even if social contexts are not *the* best way to maximize truth, developing effective skills at social deliberation is surely *an* effective way to do so. Mercier and Sperber give us added reason to take seriously the goal of researching and implementing effective strategies for cultivating and practicing skills at reasoning in a social context.

Conclusion

Despite these concerns, this is an important book, and one that will guide further research in the philosophy and psychology of reasoning, as well in critical reasoning pedagogy. It is easy to forget that, in teaching critical reasoning, the notion of 'critical reasoning' itself is in dispute. Not only will finding clarity on this notion help to better justify the philosophical underpinnings of such an education, but it will have practical consequences for how we teach it. Mercier and Sperber's book, while far from definitive, pushes the discussion forward in a way that will ultimately help us to a better sense of what reason is, and how we can improve our ability to reason.

References

Beck, J. (2018). Do nonhuman animals have a language of thought? In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 46-55). New York, NY: Routledge.

Burge, T. (2010). Steps toward origins of

propositional thought. *Disputatio*, *4*(29), 39–67.
Call, J. (2004). Inferences about the location of food in the great apes (pan paniscus, pan troglodytes, gorilla gorilla, and pongo pygmaeus). *Journal of Comparative Psychology, 118*, 232–241. doi:10.1037/0735-7036.118.2.232
Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276. doi:10.1016/0010-0277(89)90023-1
Devitt, M. (2006). *Ignorance of language*. Oxford, UK: Oxford University Press.
Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20–29. doi:10.1111/j.1745-6916.2008.00058.x
Goldstein, D.G. & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75-90. doi:10.1037/0033-295X.109.1.75
Maynes, J. (2013). Thinking about critical thinking. *Teaching Philosophy*, *36*, 337–351. doi:10.5840/teachphil2013931
Maynes, J. (2015). Critical thinking and cognitive bias. *Informal Logic*, *35*, 183–203. doi:10.22329/il.v35i2.4187
Maynes, J. (2017). Steering into the skid: On the norms of critical thinking. *Informal Logic*, *37*, 114–128. doi:10.22329/il.v37i2.4818
Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48. doi:10.1016/j.cognition.2016.05.012
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259. doi:10.1037/0033-295X.84.3.231
Proust, J. (2018). Non-human metacognition. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 142-154). New York, NY: Routledge.
Shanahan, M. (2016). The frame problem. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2016 ed.). Retrieved from https://plato.stanford.edu/entries/frame- problem/
Willingham, D. T. (2007). Critical thinking. *American Educator*, *31*, 8–19.

Author Information

Jeffrey Maynes is an Associate Professor of Philosophy at St. Lawrence University, working in the philosophy of mind and language, as well as the theory and practice of critical thinking pedagogy. He can be reached at jmaynes@stlawu.edu.