Review

Hugo Mercier and Dan Sperber, *The Enigma of Reason* (Cambridge, MA: Harvard University Press, 2017). 396 pp. \$29.95 (hbk). ISBN: 978-0-67436-830-9.

Reviewed by: Bryon Cunningham, California School of Professional Psychology, Alliant International University, Los Angeles, CA, USA
bryoncunningham@gmail.com

The Enigma of Reason begins, as the title suggests, with a puzzle. Or more precisely, two puzzles. The first puzzle is that human reason appears to be a *unique* cognitive capacity among living things. The second puzzle is that human reason appears to be a *flawed* cognitive capacity in a variety of ways. Hugo Mercier and Dan Sperber set out to solve both puzzles with a new theory of reason. Expanding and refining their earlier work (Mercier and Sperber 2011), their new theory offers a fresh perspective on what reason *is*, what it is *for*, and how it *works*.

What reason is: Mercier and Sperber argue that reason is a specialized mental module that produces intuitive inferences about reasons (i.e. facts that support a conclusion). This is a departure from traditional theories of reason in nearly every respect. By characterizing reason as a *specialized mental module*, they depart from the traditional view that reason is a *general-purpose mental faculty*. By characterizing reasoning as *intuitive*, they depart from the traditional view that the operation of reason is *consciously accessible*. Even their characterization of reasoning as *inferential* is a departure from the traditional view that reasoning is a process of following logical rules or principles (of deduction, induction, abduction, and so on). Instead, they compare the inferences of reason to the “inferences” of perception, memory, and language comprehension – inferences that are autonomous and domain-specific. Altogether, Mercier and Sperber offer a radical reconceptualization of the nature of reason.

What reason is for: adopting what they call an “interactionist” perspective, Mercier and Sperber argue that the primary function of reason is to

enhance *social coordination and communication* through justifications and arguments. This is a repudiation of conventional “intellectualist” theories according to which the primary function of reason is to enhance *the cognition of individuals* for processes like belief formation and decision making. Mercier and Sperber review a range of experimental evidence that shows that human reasoning is often illogical, lazy, biased, obstinate, overconfident, self-serving, and unconsciously motivated. These now widely-recognized results have been commonly interpreted as evidence that human reason is a poorly designed system that inadequately fulfils its primary function. But, as Mercier and Sperber point out, another conclusion is possible, namely that the primary function of reason has been generally misunderstood. Adopting a view now common in evolutionary psychology, the authors argue that the primary function of reason should be identified by asking why reason evolved in ancestral environments. Their answer is that reason evolved to enhance social coordination and communication through the production and evaluation of justifications and arguments. According to Mercier and Sperber, the *production of justifications* helps people protect and enhance their reputations within groups, and the *evaluation of justifications* helps people establish common expectations and commitments within groups. The *production of arguments* helps people persuade sceptical group members to adopt a belief or behaviour, and the *evaluation of arguments* helps people maintain “epistemic vigilance” so as to avoid being duped by cheaters, liars, and fools.

How reason works: Mercier and Sperber argue that reason is a specialized mental module whose *inputs* are mental representations of facts, whose *operation* is a process of intuitive inference, whose *outputs* are intuitive inferences about reasons, and whose *proper domain of functioning* is interpersonal interaction. Taking them one at a time:

- a. *The inputs of reason:* Mercier and Sperber argue that the inputs of the reason module are *mental representations of facts*. The authors are careful to distinguish mental representations of facts from *psychological reasons*, which are mental representations of facts that support a conclusion. They argue that mental representations of facts, but *not* psychological reasons, are the standard inputs of the reason module. They assert that “[m]odules [...] 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” (p. 120)¹. This is a rejection of the role that psychological reasons play in conventional intellectualist theories of reason, namely the role of motivating and guiding beliefs and behaviour.

- b. *The operation of reason*: Mercier and Sperber argue that the operation of the reason module is a process of *intuitive inference*. The authors intend “intuitive” to be understood not as a descriptor of a distinct mental faculty, but rather as a descriptor of any mental process whose output is conscious but whose operation is not. Against popular wisdom, they argue that reasoning itself is an intuitive mental process. This contradicts recent dual-process theories according to which intuition and reasoning are distinct cognitive systems for “thinking fast” and “thinking slow”, respectively (e.g. Kahneman 2011). Mercier and Sperber insist that “reasoning is not an alternative to intuitive inference; reasoning is a use of intuitive inferences about reasons” (p. 133). The authors intend “inference” to be understood in the nominal sense of extracting novel information from available information, in the way that object boundaries are “inferred” from shading in visual perception. This contradicts the traditional view that reasoning is a process of following logical rules or principles (of deduction, induction, abduction, and so on), and replaces it with the view that reasoning is an autonomous, domain-specific procedure for producing intuitive inferences about reasons.
- c. *The outputs of reason*: Mercier and Sperber argue that the outputs of the reason module are *intuitive inferences about reasons*, i.e. second-order mental representations about facts that support a conclusion. Notice again that, according to the authors, reasons are not the typical inputs of the reason module – they are the typical outputs. Moreover, they argue that the reasons produced by the reason module are often not the *actual causes* of beliefs and behaviour, but are merely *rationalizations* invoked to defend oneself or to persuade others, often by showing how a belief or behaviour is consistent with established social norms. As the authors put it, “[h]umans are rationalization machines” (p. 253). While that contradicts the romanticized view of reason as a paragon among human faculties, Mercier and Sperber contend that it is much more consistent with the empirical evidence about how people actually reason.
- d. *The proper domain of reason*: Mercier and Sperber argue that, for reason to perform at its best, *people must interact with one another*. They assert that “[t]he proper domain of reasoning is disagreements between oneself and others – clashes of ideas. Reasoning aims at reducing these disagreements by producing arguments to convince others or by evaluating the arguments of others and possibly be convinced oneself” (p. 289). Mercier and Sperber suggest that, when people *produce* justifications and arguments, they are biased and

lazy, but when they *evaluate* justifications and arguments, they are objective and demanding. This asymmetry, they suggest, results in an efficient distribution of cognitive labour within groups of interacting individuals. That efficient distribution of cognitive labour, they say, accounts for the fact that groups perform better than individuals on many experimental reasoning tasks like the Wason 4-card selection task. The efficient distribution of cognitive labour also plays an important role, they propose, in jury deliberation, moral reasoning, education, scientific research, and democratic traditions.

At the heart of Mercier and Sperber's *The Enigma of Reason* is their explanation of the successes and failures of human reasoning – “the bright side” and “the dark side” of reason, as they call them. Their explanatory approach is itself a blend of two styles of reasoning that have received much recent attention from philosophers of science, namely adaptationism and inference-to-the-best-explanation. Adaptationism suggests that the best way to understand the nature of many biological and psychological traits is to identify their function in the ancestral environments in which they evolved by natural selection. Inference-to-the-best-explanation suggests that the fact that a hypothesis can successfully explain a phenomenon is at least preliminary evidence that the hypothesis is true.

How readers will react to Mercier and Sperber's theory of reason will be largely determined by their attitudes toward adaptationism and inference-to-the-best-explanation. Proponents of adaptationism may be sympathetic to Mercier and Sperber's hypothesis that the evolved function of reason is to enhance social coordination and communication through justifications and arguments. Opponents of adaptationism are likely to point out that this hypothesis is difficult or impossible to test directly, or that the shortcomings of human reason are better explained by computational constraints than by natural selection. Proponents of inference-to-the-best-explanation may be sympathetic to Mercier and Sperber's hypothesis that reason is a specialized mental module that produces intuitive inferences about reasons. Opponents of inference-to-the-best-explanation are likely to point out that other theories can also explain the observed patterns of human reasoning, and that the existence of a reason module is largely speculative. Mercier and Sperber acknowledge as much when they admit, “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” (p. 144).

Another central postulate of Mercier and Sperber's theory that remains speculative is their prediction that people are much more objective and demanding in their *evaluation* of justifications and arguments than they

are in the *production* of them. As they point out, “[t]here is hardly any direct evidence on the issue in the literature, and the little there is is inconclusive” (p. 332). The speculative nature of Mercier and Sperber’s theory can be felt more generally in their liberal use of fictional examples, just-so stories, and intuition pumps. In a nutshell, Mercier and Sperber’s success at *abducting interesting hypotheses* from the research on human reasoning surpasses their success at *adducing compelling evidence* that those hypotheses are correct. But given that scientific study of human reasoning is less than 60 years old (measured from the publication of Wason’s first major finding on confirmation bias in 1960), this is a forgivable sin of omission.

On the other hand, a rather puzzling omission is the absence of any discussion of research on *distributed cognition*. Mercier and Sperber emphasize throughout the book that “reason didn’t evolve to enhance thinking on one’s own but as a tool for social interaction” and that “people can find together solutions that none of them could find individually” (p. 333). This self-described “interactionist” perspective shares much in common with the theoretical orientation of researchers working on distributed cognition (e.g. Zhang and Patel 2006). Mercier and Sperber are of course not obligated to adopt that framework, but it would have been helpful for them to discuss how they view the relation between their theory and that growing body of research. Similarly, they dismiss the possibility that reason is a group-level adaptation in the final two pages of the book, which seems a bit slapdash in light of their heavy reliance on adaptationist arguments and their repeated emphasis on group-level interactions.

The Enigma of Reason is an ambitious and thought-provoking contribution to research on human reason. It questions many of the assumptions that underlie traditional theories of reason, and it challenges currently popular views about what reason is, what it is for, and how it works. The book is likely to draw fire both from proponents of dual-process theories of reason and from advocates of traditional intellectualist approaches to reason. The outcome of that likely debate is difficult to predict, as is the possibility of the future discovery of a “reason module”. Ultimately, these are questions to be settled not by theoretical speculation, but by empirical investigation. Nevertheless, in this reviewer’s opinion, Mercier and Sperber’s interactionist theory is an important step in the interactive process of developing a mature theory of human reason.

Endnote

1. Unspecified parenthetical referencing is from Mercier and Sperber’s *The Enigma of Reason*.

References

- Mercier, H., and D. Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory". *Behavioral and Brain Sciences* 34(2): 57–74. <https://www.doi.org/10.1017/S0140525X10000968>
- Kahneman, D. 2011. *Thinking Fast and Thinking Slow*. New York: Farrar, Straus & Giroux.
- Zhang, J., and V. L. Patel. 2006. "Distributed Cognition, Representation, and Affordance". *Pragmatics & Cognition* 14(2): 333–41. <https://doi.org/10.1075/pc.14.2.12zha>