teorema
Vol. XXXVIII/1, 2019, pp. 69-76
ISNN 0210-1602
[BIBLID 0210-1602 (2019) 38:1; pp. 69-76

Précis of *The Enigma of Reason*

Hugo Mercier and Dan Sperber

The Enigma of Reason opens up with a double enigma. Many scholars throughout history have thought of reason as a cognitive silver bullet, which would allow humans to innovate, to overcome their cognitive and emotional failings, to solve a wide variety of problems, and to better understand the world around them. The first enigma, then, is why only humans would be endowed with such a superpower? Why wouldn't such a capacity, with its multiple advantages, have evolved in many other organisms? The second enigma stems from the mismatch between this lofty view of reason, and reality: experience and experiments have shown time and again that human reason is as flawed, biased, and prone to mistakes as the rest of our cognition.

We offer a twofold answer to these two enigmas. First, we develop a new understanding of what reason *is*, and new hypotheses about what reason *is for* (Parts II and III). Second, we defend these hypotheses, focusing on the functional hypotheses developed in Part III (Parts IV and V).

Part I, "Shaking Dogma," offers a brief critical look at state of the art in the study of human reason. Since its inception in the 1960s, the psychology of reasoning has been mired in deep debates — about what reason is, how it works, what are the best tools to model reason, what role normative considerations should play in its study, and, above all, endless debates around a few experimental paradigms such as Wason's selection task or Kahneman and Tversky's conjunction fallacy. If debates are a healthy part of science, the lack of resolution after several decades of work, even when it comes to the interpretation of simple, well-studied problems, suggests that something quite fundamental is wrong with the way reason has been understood so far. The rise of dual process theories over the past 20 years seemed to promise a unified framework allowing different perspective would converge. However, if dual process theories have proven a very successful export — they are now popular in many

fields, from moral reasoning to behavioral economics — they are, ironically, in retreat within the psychology of reasoning itself [e.g. Bago & De Neys (2017); Melnikoff & Bargh (2018)]. Thus, in spite of many brilliant experiments and theoretical insights garnered over the past few decades, we believe the study of human reason is ripe for some serious rethinking, starting from different, and hopefully sturdier, foundations.

One of the strengths of our perspective is to precisely situate reason in relation with other cognitive mechanisms. To do so, Part II, "Understanding Inference," develops a taxonomy of cognitive mechanisms, summarized in FIGURE 1. The circles represent increasingly narrow categories of cognitive mechanisms. The fact that reason is in the center is a mere accident of the fact that reason is the focus of the current work, and an equivalent diagram could be drawn for any other cognitive mechanism.

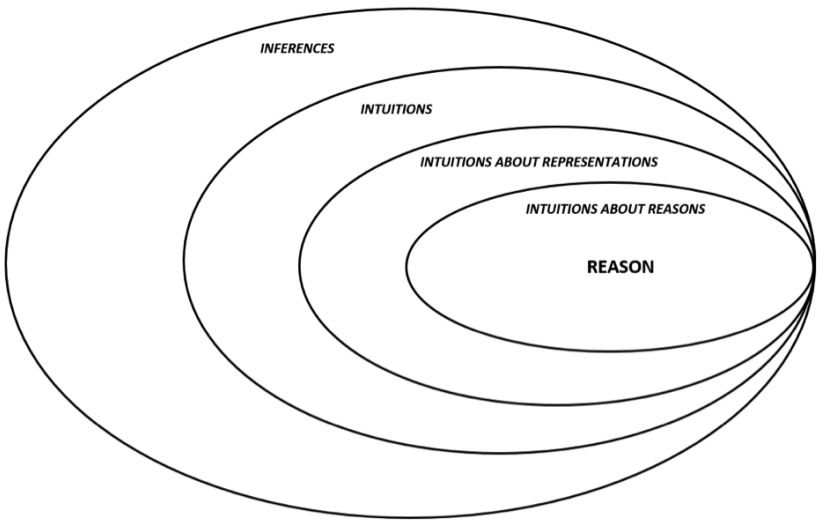

FIGURE 1. Categories of mental mechanisms
(FIGURE 1 of *The Enigma of Reason*)

The broadest category in FIGURE 1 is that of inference. Humans and other animals constantly perform inferences, building on what they already know to draw new conclusions. These inferences allow us to make sense of our perceptual environment, to anticipate the consequences of our actions, to predict what other people will do, to understand what

people mean when they speak. By contrast with some dual-process perspectives, we do not believe that humans, or any other animal for that matter, draw most of their inferences by means of some general inferential ability. Instead, humans use a wide variety of specialized inferential mechanisms, each dedicated to helping us solve a distinct problem: At what distance is an object in the visual scene? How much strength will be needed to lift some object? Is this edible? Who would make the most suitable mate? Is this animal dangerous? Fight or flee? How to react to this signal emitted by a conspecific? And so on and so forth. These specialized inferential abilities are partly instinctual, and partly the result of development and learning, with the balance varying from one ability to the next.

Plausibly all inferences, in non-human animals, and the vast majority of inferences in humans, never reach awareness — think low-level visual or syntactic processing, motor control, etc. We define intuitions —or intuitive inference — as those inferences which happen to have a conscious output together with specific metacognitive properties, such as a characteristic feeling of self-confidence [as discussed in Thompson (2014)]. For example, in a restaurant you might glance at a dish on another table and have an intuition that you'll love it. Even if you have no idea why or how you came to this conclusion, you feel as if the internal, hidden processes that led you to draw it were sound enough.

The next category depicted in FIGURE 1 is that of intuitions about representations — or metarepresentational inferences. Reason isn't the only human peculiarity. Humans are also the only animal able to fluently represent representations, and indeed humans spend much of their time doing so. Whenever we're around other people, or if we merely think of them, we represent their mental states: what their beliefs, intentions, desires are. When we engage in actual or imagined communication, we process several layers of representations [Sperber & Wilson (1995)].

Our first central claim is that reason is 'just' another type of inference. More specifically, reason is a mechanism of metarepresentational intuitive inferences that bears on one type of representations: reasons.

This claim is developed in Part III, "Rethinking Reason," in which we suggest that the study of reason (as a cognitive ability) and reasons (justifications, arguments) should be unified — even though they have often been treated separately. Reasons are essentially social. We produce reasons to justify our actions or beliefs, and to convince others. We evaluate others' reasons to decide how much belief revision they warrant. If we can use reasons on our own, in solitary ratiocination, this is only a de-

rivative usage. Moreover, in this view, the study of reason is quite divorced from that of logic (or of any other system of formal rules). Instead of being the foundation of reason, logic is a rhetorical tool that helps us express arguments more clearly by highlighting and often exaggerating the relation between premises and conclusion.

If reason isn't guided by logic, if using reason on our own is a mere byproduct, then what is the function of reason? Our second central claim is that reason mostly serves social functions, more specifically, to exchange justifications, and to exchange arguments.

More than any other primate, humans cooperate, not only with kin, but also with non-kin. This cooperation would not be sustainable if we didn't evaluate other people's reliability as partners in cooperation and in particular their competence and fairness [Baumard, André, & Sperber, (2013)]. Direct evidence of reliability is provided by the way people behave, but this evidence is limited to one's observations and is typically open to a variety of interpretations. So, we rely massively on evidence provided by others through testimony and gossip. This social evidence, when aggregated, determines a person's reputation. Having a good reputation as a reliable cooperator is a condition of social success and, beyond that, of biological fitness. This gives us a strong incentive to try to protect and improve our reputation by explaining and justifying ourselves. As observers, we have an incentive to evaluate and possibly challenge these self-justifications. Producing justifications and evaluating them is, we claim, one of the two functions of reason.

Another striking feature of humans is the extent to which they communicate. Like cooperation, communication brings its own evolutionary issues: how to avoid being lied to, misled, manipulated. To these ends, we calibrate our trust in others, being more likely to take the word of someone deemed competent and benevolent [Mercier (2017); Sperber et al., (2010)]. However, by relying purely on trust, we are bound to miss out on information that is valuable but provided by a speaker we don't trust quite enough to change our minds. In this situation, the speaker can produce arguments that, unlike mere testimony, are evaluable on their own merit, independently of the source. Accepting such arguments may lead us to change our mind when trust in the source would not have been strong enough for us to do so. Such exchanges of arguments allow for a much finer grained discrimination between valuable, and inaccurate or harmful messages.

In the past, we had focused on this latter function of reason, developing an argumentative theory of reasoning [Mercier (2016b); Mercier &

Sperber (2011)]. Since, in *The Enigma of Reason*, we claim that reason also serves another social function — namely justification — we prefer to call our current approach to reason *interactionist*.

Part IV, "What Reason Can and Cannot Do," reviews evidence pertaining to the functioning of human reason and argues that the evidence is more coherent with our interactionist account, than with other perspectives on reason. If the interactionist account is correct, we should expect reason to exhibit certain features (summarized in TABLE 1). Given that reason serves very different roles when it produces reasons aimed at others, than when it evaluates reasons others aim at us, we should expect reason to exhibit different traits when it serves each of these roles.

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

TABLE 1. The main characteristics of human reason
(FIGURE 19 of *The Enigma of Reason*)

When we produce reasons, we should be heavily biased towards our point of view. We're not going to appear more rational by providing reasons why what we did was stupid; we're not going to convince someone by giving them arguments for their point of view or against ours. This explains an otherwise puzzling feature of reason: the myside bias [or confirmation bias, see Mercier (2016a)].

We might also expect that reason would be able to produce very strong reasons, to better defend ourselves and convince others. However, finding strong reasons is a cognitively demanding task, one that is best accomplished not through brute force, but by paying attention to feedback. In our account, reason evolved as it was used in social, dialogic contexts. In such contexts, if someone isn't convinced by a reason, the natural reaction is for them to provide another reason in turn — typically, a counter-argument. It is then possible for speakers to rely on the

feedback provided by their interlocutors in order to provide better reasons, rather than having to find such reasons purely on their own. As a result, we expect the production of reasons to be lazy, but in a smart way: the first reasons we muster can be pretty banal, but we should be able to adapt to the counter-arguments raised by our interlocutors.

The way we evaluate others' reasons — at least when the reasons aim at changing our mind — should be the opposite of the way we produce reasons. We should be able to recognize good reasons, even if they challenge our prior beliefs and come from sources we do not completely trust. Indeed, doing so is, according to us, the very point of reason. And we should be able to reject weak reasons, so as not to be convinced when doing so is not warranted.

These features of reason have several implications. When people reason on their own, they should mostly produce reasons to defend their pre-existing opinions. This explains why a solitary reasoner generally fails to correct their own mistaken intuitions: as reasons pile up to support these intuitions, the reasoner even risks becoming more confident or more polarized. When a solitary reasoner doesn't have a strong intuition to begin with, reason pushes them towards the decision that is most easily justified — whether it is an otherwise good decision or not (a phenomenon known as reason-based choice in the judgment and decision-making literature).

By contrast, when people reason with each other, they take turn producing reasons and evaluating others' reasons. Under the right conditions — a small group, with some common incentives but yet disagreeing over some point — discussion and the exchange of reasons works wonders. Those who have the best ideas, or some valuable insight into a problem, can convince the other group members, leading to improvements in performance, sometimes huge ones [for a review of recent evidence, see Mercier (2016b); see also, Claidière, Trouche, & Mercier, (2017)]. That the exchange of reasons allows good ideas to spread, and performance to increase, has been observed with a wide variety of contents, from logical problems to forecasting or medical decisions.

While Part IV mostly rests on evidence gathered in the lab, Part V, "Reason in the Wild," tackles similar issues, but focusing on other sources of evidence — from anthropology, history, sociology. We start by looking at the cross-culturally robustness of the main features of reason. Is the production of reason biased and lazy in different cultures? Are people everywhere able to make the best of group discussions? By and large, the answer is yes. We then turn to two domains in which rea-

son has been claimed to function in ways that are at odds with our approach. In the domain of morality, we attempt to refute a pessimistic view according to which moral arguments would be essentially inert, showing instead that good reasons can change people's minds, even on moral and emotional matters. We then turn to science, often seen as a domain of solitary geniuses figuring out grand theories on their own and failing to convince their lesser colleagues of their brilliant insights — quite the opposite of what we predict. However, work on the history and sociology of science paints a different picture, one of scientists in constant exchange (or at least revisiting past exchanges and anticipating new ones), who push each other to develop better arguments, and whose theories can rapidly take over a field, as soon as they are supported well enough.

In this book, we thus argue for an original solution to the two enigmas of reason. Reason is uniquely human because it evolved in response to selection pressures uniquely faced by humans — cooperation and communication of an unrivalled scale and complexity. These selection pressures gave rise to inferential mechanisms dedicated to the processing of reasons for social consumption: to exchange arguments and justifications. Reason is not a superpower, it is just another specialized cognitive mechanism, whose strengths and weaknesses can be explained in the light of its function, and that may do wonders but at the social and cultural scale rather than at the individual scale.

Institut Jean Nicod, Département d'études cognitives,
ENS, EHESS, PSL University, CNRS,
Paris France
E-mail: hugo.mercier@gmail.com

Department of cognitive science and Department of philosophy,
Central European University, Budapest, Hungary
Institut Jean Nicod, Département d'études cognitives,
ENS, EHESS, PSL University, CNRS,
Paris France
E-mail: dan.sperber@gmail.com

ACKNOWLEDGMENTS

HUGO MERCIER'S work is supported by the Agence Nationale de la Recherche, EUR FrontCog ANR-17-EURE-0017. DAN SPERBER'S work is supported by the European Research Council under the European Union's Seventh

Framework Programme (FP7/2007-2013)/ERC grant agreement n° [609819], SOMICS.

REFERENCES

BAGO, B., & DE NEYS, W. (2017), "Fast Logic?: Examining the Time Course Assumption of Dual Process Theory"; *Cognition*, *158*, pp. 90-109.

BAUMARD, N., ANDRÉ, J. B., & SPERBER, D. (2013), "A Mutualistic Approach to Morality: the Evolution of Fairness by Partner Choice"; *Behavioral and Brain Sciences*, *36*(01), pp. 59-78.

CLAIDIÈRE, N., TROUCHE, E., & MERCIER, H. (2017), "Argumentation and the Diffusion of Counter-Intuitive Beliefs"; *Journal of Experimental Psychology: General*, *146*(7), pp. 1052-1066.

MELNIKOFF, D. E., & BARGH, J. A. (2018), "The Mythical Number Two"; *Trends in Cognitive Sciences*, *22*(4), pp. 280-293.

MERCIER, H. (2016a), "Confirmation (or Myside) Bias"; In R. Pohl (Ed.), *Cognitive Illusions* (2nd ed., pp. 99-114). London: Psychology Press.

— (2016b), "The Argumentative Theory: Predictions and Empirical Evidence"; *Trends in Cognitive Sciences*, *20*(9), pp. 689-700.

— (2017), "How Gullible Are We? A Review of the Evidence from Psychology and Social Science"; *Review of General Psychology*, *21*(2), p. 103.

MERCIER, H., & SPERBER, D. (2011), "Why Do Humans Reason? Arguments for an Argumentative Theory"; *Behavioral and Brain Sciences*, *34*(2), pp. 57-74.

SPERBER, D., CLÉMENT, F., HEINTZ, C., MASCARO, O., MERCIER, H., ORIGGI, G., & WILSON, D. (2010), "Epistemic Vigilance"; *Mind and Language*, *25*(4), pp. 359-393.

SPERBER, D., & WILSON, D. (1995), *Relevance: Communication and cognition*. New York: Wiley-Blackwell.

THOMPSON, V. A. (2014), "What Intuitions Are... and Are Not."; In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 60, pp. 35-75). Burlington: Academic Press.