

Hugo Mercier and Dan Sperber: *The Enigma of Reason*  
 Harvard University Press, Cambridge (Mass.), 2017, 396 pages<sup>1</sup>

At one point in their book, Mercier and Sperber present their readers with what looks like a defective chair. However, the authors point out, it looks so obviously defective only till we realize that it is not a chair, but something different, namely a kneeler. And Mercier and Sperber are suggesting that an analogous obstacle has been hampering our assessment of human *reason* – we have been brooding over its apparent shortcomings, or defectiveness, because we have been misconstruing its principal function.

The term *function* of course, by itself, is potentially ambiguous, but Mercier and Sperber are using it in its well defined sense tied to the context of evolution theory; here the function of an organ or an ability of an organism is what this organ or ability has been *selected for*. From this viewpoint, reason and reasoning is usually thought about as an adaptation helping us to solve problems, to accumulate faithful knowledge of the world and to peruse it in a cooperative way.<sup>2</sup> And given this, we humans should be expert reasoners, making errors only when facing problems that are overly complex, or when trying to solve them under significant stress. How come, then, that as a matter of fact, we make systematic errors when solving some *prima facie* simple tasks, such as the Wason task?

There are various ways to explain such spectacular failures of human reasoning. One way is by appealing to the concept of bounded rationality (see, e.g., Morton 2010): human reason is powerful, but not almighty. Failures are to be expected; we should not measure the performances of reason by abstract standards which do not take into account human limits, such as the restricted capacity of human memory or its limited and not completely robust computational powers. But the pointed question remains as to whether this is adequate to fully explain why humans can sometimes predictably fail to solve problems arguably much simpler than those which they can solve easily.

The idea proposed by Mercier and Sperber is that in fact reason is not at heart the kind of adaptation that has usually been assumed. Rather, it has been formed

---

<sup>1</sup> ✉ Jaroslav Peregrin  
 Institute of Philosophy, Czech Academy of Sciences  
 Jilská 1, 110 00 Praha 1, Czech Republic  
 e-mail: peregrin@flu.cas.cz

<sup>2</sup> Cf., for example, the notion of *cognitive niche* of Pinker (2010).

by selection pressures different from those centred on favoring the most perfect solutions of problems:

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. (p. 121)

How could this be? Is reason not a tool for solving problems? It is quite clear that we do use it to solve problems, and in many cases with superb effects; however, their claim is that this is not the function it has from the viewpoint of evolution, which would make sense of its seemingly unexplainable failures in some simple cases. What, then, would reason and reasoning have been selected for?

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 ourselves are best guided by our own intuitions (which are based, we are sure, on good reasons, even if we cannot spell them out). (pp. 122-123)

Hence what Mercier and Sperber are suggesting is that reasoning did not originally come into being as a means of solving problems, but rather as a means of coping with each other within human communities, in particular, as a means of explaining and justifying oneself. As a result, we tend to excel at producing reasons which are impressive and persuasive without necessarily being fully sound; and we also tend to excel at checking the reasons of others for their soundness.

I think this grand picture is extremely interesting and offers us a fresh vista on reason and reasoning. The book also offers a very clear explanation of this picture and it is well documented with background material. Moreover, Mercier and Sperber fill in a lot of the picture's details; some of them again quite novel and interesting; others raising some doubts.

On the ground level, Mercier and Sperber see human minds as essentially modular, where each of the modules does its cognitive work largely independently of others and using nothing like reasoning:

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. (p. 129)

Thus, though the popularity of the modular theory of mind appears to be declining (cf., e.g., Prinz 2006), Mercier and Sperber still use it as the point of departure for their theory of reasoning. What then, more precisely, is the cognitive work that the individual modules do? The short answer given by Mercier and Sperber is the drawing of *inferences*.

This is one of the points which I find puzzling. It is clear that on the modular theory of mind, each module takes care of coping with some part or aspect of the world; but why should this always be a matter of inference? If I understand the authors properly, their answer is contained in the following assumption:

A main goal of cognitive mechanisms is to maintain an accurate representation of the organism's environment, or at least of relevant aspects of it. (p. 218)

An inference, then, is a mechanism that enables the module to produce further representations which can be used for prediction. However, if the main goal of cognitive mechanisms is to cope with the organism's environment (which, to be sure, *may sometimes* – or perhaps *often* – be achieved by manipulating representations of the environment), then inferring does not seem to be the common core of the modules' functioning. (Consider, for example, the theories of situated cognition, going back to Brooks (1991) and others: according to these, lots of coping with the environment can be done wholly avoiding representing the environment and manipulating its representations.)

An important question, of course, is what exactly *is* inference? The authors, for example, speak about inferences inherent to our visual perception, which seems to indicate that inferring is not necessarily something which we *do* with our representations, it may be something that *happens* to us. Then of course, it is less difficult to squeeze anything what the modules do into the boxes of *representing* and *inferring*, but then the concepts would not seem to be very useful.

Anyway, reason, on Mercier's and Sperber's view, turns out to be merely one new module. It is a module which, in effect, reflects on some of our inferences and seeks what makes us draw them, what we see as the (real or alleged) *reasons* for their outcomes. It seeks them especially because we might need them to negotiate our position within our society:

We show, in other terms, how reason fits among other modules of intuitive inference rather than being a towering superpower. Notwithstanding its virtual domain generality, reason is not a broaduse adaptation that would be advantageous to all kinds of animal species. Reasons, we argued, are for social consumption. Reason is an adaptation to the hypersocial niche humans have built for themselves. (p. 339)

However, once we accept that we became inferring creatures without becoming reasoning creatures, the story that takes us to reasoning proceeds quite smoothly. It is, in essence, a story about us coming to reflect upon our inferences. In this way, we come to reflect that we have certain representations or do certain things because we came to have other representations, and we construct the picture of our peers – and of ourselves as *acting for reasons*:

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. (p. 136)

Reasoning thus is primarily tied to social contexts and to language – in its primordial shape it is argumentation (and originally not even argumentation in the sense of cooperatively finding an objective truth, but in the sense of competitively negotiating one's position in a society). Reasoning as an inner mental process is parasitic on argumentation – thus it is also an essentially linguistic matter:

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. (p. 172)

Is the upshot, therefore, that it is merely illusory to believe that reasoning is an extremely useful tool that helps us attain knowledge and solve problems with a sophistication far beyond the ken of animals unable to reason? Is reason merely an

advocate that seeks to find justification for our preconceptions, disguised as an impartial judge seeking the truth? This is not quite the message of the book. Mercier and Sperber agree that reason can lead us to valuable conclusions, only it must be used in the proper way, where using it in this proper way means using it so that it chimes with its primordial function as much as possible.

What should we do if we want to reason with such beneficial effects? The most important thing, Mercier and Sperber argue, is that we should reason interactively:

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 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. (pp. 207-208)

We are ingenious in coming up with reasons, though they are not always entirely sound. But we are also ingenious at checking the reasons of other people for their soundness, so when we work in coordination, opposing one another's tendency to spout not always very good reasons (or "reasons"), the interaction may yield something not so far from impartial reasoning and homing in on objective truth.

And what holds about reasoning in general, holds equally about what is often thought about as the quintessence of reasoning, science:

Scientists' reasoning is not different in kind from that of laypeople. Science doesn't work by recruiting a special breed of superreasoners but by making the best of reasoning's strengths: fostering discussions, providing people with tools to argue, giving them the latitude to change their minds. (p. 329)

What, however, the view of reason and reasoning put forward by Mercier and Sperber does shatter is the recently popular view of man as a *Homo economicus*, who consistently maximizes his gains by means of his reason, which has developed precisely to serve this purpose. This theory, in contrast to Mercier's and Sperber's, sees reason as primarily an individual adaptation for finding objectively best solutions to problems the individual faces.

But does it not follow directly from evolution theory that surviving animals must become experts in solving the problems they encounter? And if so, should this not form also our reason – our principal means of the coping? Their answer is that being the kind of (hyper)social animals we humans are, most of the existential problems that we solve we face *as a society*, and the ability to negotiate one's position in a society is even more important for an individual than to directly handle natural menaces. Though it is probable that reason was, perhaps from the beginning, used *also* to solve problems and to deal with the environment, its social function was so much more important that it was this that shaped it.

Again, I think that in their zeal to revert the reader from the mistaken mainstream view of reasoning, the authors sometimes make dubious claims. Thus they write:

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. (p. 166)

Well, I think that undeniably a sound syllogism does provide us with a sufficient argument; or if we doubt its premises, it moves us a step towards such an argument. What the authors probably want to say is that the syllogism often *de facto* does not serve as such an argument, that it can be used and misused in various ways and that we may have alternative means of becoming convinced of its conclusion.

On the whole I think that Mercier's and Sperber's book is extremely interesting and duly thought-provoking. And I suspect that their view of reason and reasoning, path breaking as it is, is largely correct – perhaps not in all details, but surely in the general outline. And I think that its consequences for our studying reason and reasoning are huge.

*Jaroslav Peregrin*

### References

- BROOKS, R. A. (1991): Intelligence without Representation. *Artificial Intelligence* 47, 139-159.
- MORTON, A. (2010): Human Bounds: Rationality for Our Species. *Synthese* 176, No. 1, 5-21.
- PINKER, S. (2010): The Cognitive Niche: Coevolution of Intelligence, Sociality, and Language. *PNAS* 107, 8993-8999.
- PRINZ, J. (2006): Is the Mind Really Modular? In: Stainton, R. J. (ed.): *Contemporary Debates in Cognitive Science*. Oxford: Blackwell, 22-36.