ROUTLEDGE INTERNATIONAL HANDBOOKS

Routledge Handbook of Bounded Rationality

Edited by Riccardo Viale

16

BOUNDED REASON IN A SOCIAL WORLD

Hugo Mercier and Dan Sperber

Introduction

Herbert Simon (1983, pp. 34–35) distinguished three "visions of rationality": (1) the "Olympian model," which "serves, perhaps, as a model of the mind of God, but certainly not as a model of the mind of man;" (2) the "behavioral" model, which "postulates that human rationality is very limited, very much bounded by the situation and by human computational powers;" and (3) the "intuitive model," which "is in fact a component of the behavioral theory." Bounded rationality, with its intuitive component, is to be explained, Simon adds, in an evolutionary perspective. Our joint work on reasoning and in particular our book *The Enigma of Reason* (Mercier & Sperber, 2017) describes mechanisms of intuitive inference in general and the mechanism of reason in a way that is quite consistent with Simon's defense of a "bounded rationality" approach to human reason. Like other evolutionary psychologists (in particular, Leda Cosmides and John Tooby, see Tooby & Cosmides, 1992) and like Gerd Gigerenzer's 'adaptive toolbox' approach (Gigerenzer, 2007; Gigerenzer, Todd, & ABC Research Group, 1999), we don't see bounded rationality as an inferior version of Olympian rationality, nor do we think that human or other animal inferences should be measured against abstract rationality criteria. Our distinct contribution is to argue that there is an evolved mechanism that can reasonably be called "reason," the function of which is to address problems of coordination and communication by producing and evaluating reasons used as justifications or as arguments in communicative interactions.

Other approaches to human reasoning make an important but less comprehensive use of the idea of bounded rationality. Tversky and Kahneman's heuristics and biases program (Gilovich, Griffin, & Kahneman, 2002) is a case in point. The heuristics studied by Tversky and Kahneman rely on regularities in the environment to make broadly sound decisions. For instance, everything else equal, more salient information is more likely to be relevant, making the availability heuristic sensible. However, this tradition has mostly focused on the biases and errors that result from the use of heuristics (Kruger & Savitsky, 2004), suggesting that an alternative way of processing information could lead to superior results.

The heuristics and biases program has now become largely integrated into the dual process paradigm, along with other strands of research, from social psychology to reasoning—and the dual process theory has become dominant (see Melnikoff & Bargh, 2018). In this

paradigm, the mind is split into two types of processes. System 1 processes (intuitions) are fast, effortless, unconscious, and prone to systematic mistakes. System 2 processes (reasoning) are slow, effortful, conscious, and able to correct System 1's mistakes (e.g., Evans, 2007; Kahneman, 2003; Stanovich, 2004; for a more recent take on dual process models, see, e.g., Evans & Stanovich, 2013).

By and large, System 1 behaves as would be expected by models of bounded rationality: satisficing, relying on heuristics. System 2, by contrast, is closer to the ideal rational agent, able to correct any type of mistake made by System 1, to follow normative guidelines and yield strictly rational decisions.

However, it's far from clear how System 2 could possibly be a maximizer rather than a mere satisficer. The arguments put forward by Simon in favor of a view of rationality as bounded should apply to both System 1 and System 2. Relatedly, evolutionary psychologists have pointed out that cognitive mechanisms tend to specialize, allowing them to incorporate relevant environmental regularities and function more efficiently. System 2, by contrast, would be the ultimate generalist, able to fix the mistakes made by countless different System 1 processes, from social to statistical heuristics. Unsurprisingly, then, dual process models have been plagued by conceptual difficulties. How does System 2 know when to override System 1? How is System 2 able to find the appropriate reasons to counteract each and every System 1 heuristic? (see, e.g., De Neys, 2012; Osman, 2004).

Moreover, talking, for the sake of argument, in terms of Systems 1 and 2, it is not hard to show that System 2 fails at performing a function it couldn't fairly be expected to perform. By and large, solitary reasoning doesn't correct mistaken intuitions. The large share of participants—often the majority, sometimes the overwhelming majority—that do not provide the normatively correct answer to simple decision making or reasoning problems reflect a massive System 2 failure. After all, in most cases, the answer should be easily available to the participants.

Consider the well-known bat and ball problem (Frederick, 2005):

> A bat and a ball cost $1.10 together.
> The bat costs $1 more than the ball.
> How much does the ball cost?

Most participants—college students well able to solve this trivial mathematical problem—give the intuitive but wrong answer of 10c, a blatant System 2 failure. Recent experiments have revealed an even more dire picture for standard dual process models, showing that of those participants who end up providing the correct answer (5c), the majority was able to do so immediately after seeing the problem, and so (from the point of view of dual process theory) through System 1 processes. Overall, only a few percent of the participants behaved as expected in standard dual process model, making an intuitive mistake and later correcting it thanks to System 2 (Bago & De Neys, 2019). The same pattern has been observed using other classic reasoning and decision-making problems and other methods (Bago, 2018; Bago & De Neys, 2017).

The reason why System 2, as a rule, fails to correct mistaken intuitions is even more of an indictment of standard dual process models. Instead of looking for reasons why our intuitions might be mistaken, or to look for reasons supporting alternative answers, System 2 suffers from a massive myside bias (or confirmation bias, see Mercier, 2016a). Once someone has an intuition about what the correct decision is, their System 2 mostly finds reasons supporting this intuition. Moreover, System 2 is lazy when it comes to evaluating our own reasons, not making

sure that the reasons we find to support our intuitions are particularly foolproof (Mercier, Bonnier, & Trouche, 2016; Trouche, Johansson, Hall, & Mercier, 2016). Confirmation bias and laziness were supposed to be System 1, not System 2 features, but the evidence says otherwise.

In light of these results, and of the fundamental problems that affect standard dual process models, we have suggested an alternative theory that is much more in line with the ideas of bounded rationality (Mercier & Sperber, 2011, 2017). Instead of dividing the mind between System 1—which would mostly abide by the dictates of bounded rationality—and System 2—which wouldn't—we suggest that it is intuitions all the way up (see also Kruglanski & Gigerenzer, 2011; Osman, 2004). However, some of these intuitions bear on reasons, as explained presently. In this theory, reason heavily relies on its cognitive and social environment to solve problems—when to kick in, how to figure out if something is a good reason, and how to find relevant reasons.

An interactionist view of reason

Our account—dubbed the interactionist view of reason—differs from most existing accounts of reason—in particular, standard dual process models—on two grounds: what reason is, and what reason is for.

Following the arguments put forward by evolutionary psychologists (e.g., Sperber, 1994; Tooby & Cosmides, 1992), we suggest that the mind is a collection of cognitive mechanisms, or modules, that function each in an autonomous manner and that are related to one another by input/output relationships. Many of these modules perform inferences in which they process an input and informationally enrich or transform it in an epistemically sound way. Some of these inferential processes are completely unconscious—neither the input nor the output is consciously accessed. For others, the output is conscious and is experienced as an intuition (Thomson 2014). For example, the attribution to a speaker of an ironical intent is the conscious output of intuitive mechanisms of verbal comprehension (Wilson & Sperber, 2012). As suggested by this example, some intuitions are metarepresentational: they take as input, and/or deliver as output, representations of representations (see Sperber, 2000). The two best-studied such mechanisms deal, one, with attribution of mental states such as beliefs and desire and, the other, with comprehension, that is, with the attribution of speaker's meaning.

We suggest that reason is another of these metarepresentational mechanisms. It takes several representations as inputs—the premises and conclusion of an argument—and delivers as output an intuitive metarepresentational judgment of the quality of the support relation between premises and conclusion. Crucially, this means that reason is 'just' another cognitive mechanism among many others, which shares all the typical traits of intuitions, being fast, effortless, and involving no consciousness of the process through which the outputs are arrived at. Consider an everyday argumentative discussion—about which restaurant to go to, or which printer to buy. Reasons are generated and evaluated quickly and effortlessly, and, most of the time, with no insight as to why such and such reason was chosen, or why a given reason was found to be good or bad. Sometimes we have higher-order reasons for finding a lower reason good or bad. Such higher-order insight into the strength of lower-order reasons is, however, not to be confused with an introspective access to the inferential process itself.

When we entertain higher-order reasons for our lower-order ones, or when we use a chain of reasons, this typically involves an effort of working memory. This role of working memory has sometimes be taken to be diagnostic of reasoning proper (e.g., Evans, 2003). This kind of memory effort, however, is not really different from that of some non-reasoning tasks such as

looking for one's keys left somewhere in the flat and trying to avoid both overlooking a possible place and looking twice in the same place. Unlike keeping in mind a hierarchy or a chain of reasons, the production and evaluation of each individual reason in such a hierarchy or chain are relatively effortless, as are all individual intuitive processes.

When sequences of reasons are deployed in a dialogic form, they are distributed among interlocutors, doubly diminishing the memory task: on the one hand, the individual who first invokes a given reason is motivated to bring it back into the discussion as long as it is still relevant; on the other hand, in the mind of the interlocutors, each reason is associated with its producer, providing, for remembering each individual argument, an analog of the method of loci in classical mnemonics. By contrast, in solitary reasoning these memory props are not really available. True, a solitary reasoner can imaginatively produce an internal debate of sorts. Such an ersatz is unlikely to offer the same mnemonic and dialectic benefits as a real debate. Even more relevantly here, the fact that solitary reasoning typically resorts to imagining a dialogue (just as solitary sexuality typically resorts to imagining interactive sex), underscores the second originality of our view: the claim that reason not only performs social functions (as suggested, for instance, by Billig, 1996; Gibbard, 1990; Haidt, 2001; Kuhn, 1992; Perelman & Olbrechts-Tyteca, 1958) but actually evolved to do so.

For a very long time, humans and their ancestors have evolved not only in very social environments, but more specifically in very cooperative environments, collaborating to perform a variety of tasks, from hunting to raising children. This high degree of collaboration was sustained by partner choice: people competing to be seen as reliable cooperation partners, so they would be included in more collaborations (Baumard, André, & Sperber, 2013). This means that individuals had strong incentives to maintain a good reputation as competent and diligent collaborators (Sperber & Baumard, 2012). One way of doing so is by justifying our actions. Whenever we do something that might look irrational or immoral, we can offer justifications to change the audience's mind. Figuring out whether other people's actions really are irrational or immoral involves taking into account their point of view. Hence, the audience has an incentive to listen to these justifications and, if they are found to have some merit, adapt their judgments accordingly.

Another consequence of humans' high degree of cooperation is their unprecedented ability to communicate. However, communication is a risky business, as one might be misled, manipulated, or lied to (see Maynard Smith & Harper, 2003). Humans have evolved a suite of cognitive mechanisms to limit these risks, for instance by calibrating their trust in different people as a function of their competence and diligence (Mercier, 2017, 2020; Sperber et al., 2010). However, trust calibration has limits. There are situations in which a piece of information is too important to be accepted just on trust; there are situation where communicating a piece of genuine information that could be misinterpreted may involve too much of a risk to one's reputation of trustworthiness. The best way to safeguard communication in such situations is to appeal not to authority and trust but to properties of the content of the information itself: in particular, its consistency with what is already accepted as true. Displaying this consistency is precisely what reasons do. In an exchange of arguments, speakers provide reasons to support their point of view, and their interlocutors evaluate these reasons to decide whether or not they should change their minds.

The interactionist view of reason claims that human reason is a metarepresentational cognitive mechanism, or module, used to produce reasons and to evaluate them (for a detailed account of this modularist approach, see Mercier & Sperber, 2017; for a defense against some common objections, see, Sperber & Mercier, 2018). Reason evolved because it allowed individuals to find justifications and arguments, and to evaluate the justifications and arguments

offered by others, thereby overcoming the limitation of trust and allowing better coordination and communication.

When is reason triggered?

In our model, the most basic triggers of reason are social. When it comes to evaluating others' reasons, the trigger is simply being presented with a statement as a reason for some conclusion. That a statement should be understood as a reason for a given conclusion can be made explicit (for example, with connectives) but is often left implicit. When it comes to producing reasons, the basic trigger is the expression by others of criticism or doubt about our own actions or opinions (which triggers a search for justifications) or of disagreement (which triggers a search for arguments). Again, criticism or disagreement can be expressed more or less strongly and explicitly (from "that was stupid" to a frown, or even a lack of clear agreement).

While these basic triggers may be entirely external, with experience, we learn to internalize them, in particular by anticipating the need to offer reasons. For instance, adults almost always justify actions before being questioned about them (Malle, 2004). Even young children are, to some extent, able to anticipate the need for justifications. Three- and five-year-olds are more likely to offer an explicit reason supporting an unconventional choice than a conventional one (Köymen, Rosenbaum, & Tomasello, 2014). The mechanisms that allow us to anticipate the need to offer reasons are distinct from reason itself, encompassing a variety of social cognitive mechanisms.

Looking for reasons in anticipation of having to provide them can have several consequences. If we start from a strong intuition, this anticipatory search of reasons typically yields a series of poorly examined reasons supporting our initial intuition, making us more confident or driving us towards more extreme views (Koriat, Lichtenstein, & Fischhoff, 1980; Tesser, 1978). If we start from weak or conflicting intuitions, this anticipatory search of reasons is likely to drive us toward the intuition which is easiest to justify, whether it is also the best by others' standards or not (a phenomenon known as reason-based choice, see Simonson, 1989; Shafir, Simonson, & Tversky, 1993; and Mercier & Sperber, 2011 for review).

Another consequence of our tendency to anticipate the need for reasons is to make us aware that some pieces of information we encounter may become relevant as reasons. For example, if you often discuss politics with your friends, you'll find information that supports your political views relevant as reasons. This anticipation of the need for reasons is likely the main mechanism going against our general preference for information that clashes with our priors (and which is, by definition and everything else being equal, more informative). Anticipatory uses of reason would thus be driving selective exposure (our tendency, in some contexts, to look for information that supports our beliefs, see, e.g., Smith, Fabrigar, & Norris, 2008).

A substantial amount of evidence shows that anticipatory reason is triggered, and reinforced by social cues. For example, being accountable—when participants are told they will have to publicly justify their decisions—tends to reinforce reason-based choice (e.g., Simonson, 1989).

How does reason recognize good reasons?

Once reason has been triggered by the encounter with a reason—when someone provides us with an argument to change our mind, or when we read something in the newspaper that we could remember as a reason to defend our views—its task is to evaluate how good a reason it is: does the premise effectively support the conclusion? What makes this task difficult is that reasons can be about anything: whether abortion should be forbidden, which restaurant to go

to, whether plate tectonics is an accurate theory, and so on and so forth. There cannot be general rules for what makes a good reason. The two most popular explicit means of evaluating arguments—logic and argumentation fallacies—are at best rough approximations, and at worst misleading tools (Boudry, Paglieri, & Pigliucci, 2015; Mercier, Boudry, Paglieri, & Trouche, 2017; Mercier & Sperber, 2017).

Instead, the evaluation of reasons has to be made on a case-by-case basis. To do this, reason relies neither on a general logic nor on a general probability calculus, but on a metacognitive capacity. Reason is both a metarepresentational and a metacognitive device (Sperber & Mercier, 2018). Whenever we encounter a piece of information presented as a reason, we check whether we would intuitively derive from what is offered as a reason the conclusion the reason purports to support. For example, if your colleague tells you: "I'm sure Bill is here, I saw his car in the car park not five minutes ago," you have intuitive access to the fact that you would draw the conclusion that Bill is here from his car having been seen in the car park five minutes ago. The more confident you would be of this conclusion, the stronger you intuit the reason given by your colleague to be.

It might seem, then, that reason doesn't add anything of value. After all, if you are disposed to infer that Bill is here from the information that your colleague saw his car in the car park, what benefit is there, over and above drawing this conclusion, in interpreting your colleague's utterance as a reason for this conclusion and moreover in evaluating this reason? Actually, there are several benefits. First, from the fact that your colleague saw Bill's car five minutes ago, you could have inferred many diverse conclusions, for instance, that Bill's car is not in the garage, that your colleague looked at the car park, that the car park is not empty, and so on indefinitely. By offering this fact as a reason for this particular conclusion, your colleague indicated the way in which this fact was relevant in the situation.

Second, reason allows you to check whether a reason would be good independently of whether you accept the premise as true or not. For example, you might not be convinced that your colleague is telling the truth in claiming to have seen Bill's car—he might be in bad faith or he might have confused someone else's car for Bill's—but still recognize that if what he says were true, it would be a good reason for the conclusion that Bill is here. Obviously, this ability to evaluate how good an argument would be if its premise were true is critical in science, as it allows us to tell how evidence not yet acquired would bear on various hypotheses. Third, evaluating the quality of reasons as reasons allows us to draw inferences about the speaker's competence. People who offer poor reasons don't just fail to change their interlocutor's mind, they might also be perceived as less competent—and conversely for people who offer particularly good reasons. Fourth, by using reasons, speakers commit, within limits, to what they deem to be good reasons. Someone who uses a given reason to support their ideas, but refuses a similar reason challenging their ideas, appears inconsistent. Fifth, understanding that a statement is presented as a reason helps guide our own search for reasons, as explained presently.

How does reason find reasons?

In the interactionist view, reason fulfills its function by making it possible to produce and to evaluate reasons. Reasons are produced to justify or to convince. Should we then expect reason to be able to find, from the get-go, very strong justifications and arguments, to be able to form long, sophisticated, well-formed pleas that anticipate most potential rebuttals? As psychologists (and others) have long noted, this is not what typically happens (see, e.g., Kuhn,

1991). Instead, when asked to justify their positions, people tend to produce relatively shallow, superficial reasons—the first thing that comes to mind, or close to it. Why aren't we better at finding good reasons? The interactionist approach provides a twofold answer (see Mercier et al., 2016).

First, it would be tremendously difficult for an individual to reliably anticipate on her own how effectively her reasons would sway her interlocutor (see Mercier, 2012). In most cases, good reasons are highly context-dependent. Imagine you're trying to convince a friend to go to a given restaurant. An ideal argument would take into account your friend's culinary preferences, the kind of restaurant they've recently been to, how much money they are willing to spend, how far they'd go, and so on and so forth. Anticipating which argument in favor of your choice of restaurant will prove convincing and which will be rebuffed is hard.

Second, fortunately, our interlocutors can help us find the most relevant arguments to convince them, obviating the need to do so on our own. In a typical informal discussion, the cost of having our first argument rejected is low (unless the argument is particularly dim). For instance, you might try to convince a friend by telling them "this restaurant makes great cocktails," but they reply that they don't feel like drinking tonight. Not only did you bear no costs for failing to convince them, but thanks to this information, you can narrow down your search for arguments, or offer a direct counter-argument—"they also make great virgin cocktails." The production of reasons, as we describe it, is a paradigmatic example of a satisficing process. The criterion to be reached is that of producing reasons good enough to convince one's interlocutor.

For the interactionist account, reason evolved by being used chiefly in dialogic contexts, in which we can benefit from the back and forth of discussion to refine our arguments, instead of attempting to anticipate through extraordinary computational force what the silver bullet might be. This means, however, that people are not well prepared to produce strong reasons in the absence of feedback (except in the relatively rare circumstances where such feedback can be reliably anticipated). In ordinary circumstances, people are likely to come up with the same kind of relatively shallow reasons that work well to open a discussion, rather than attempting to imagine a variety of potential counter-arguments and trying to pre-empt them. This explains why people tend to be lazy—as mentioned above—when evaluating their own reasons, and why, partly as a result, they often fail to correct their mistaken intuitions when reasoning on their own.

Besides providing us with direct feedback guiding our search for reasons, the social environment offers other opportunities for improving one's production of reasons. One simple opportunity consists in recycling reasons provided by others, and that we found to be good reasons. Pupils have been observed to pick up on the argument forms used by their classmates (Anderson et al., 2001). Participants who have been convinced to accept the logical answer to a reasoning problem (such as the bat and ball) are then able to reconstruct the argument in order to convince other participants (Claidière, Trouche, & Mercier, 2017).

Still, even if the social environment provides many learning opportunities, producing good reasons to justify oneself and convince others is harder than evaluating reasons provided by others. Individual variations are greater in the production than in the evaluation of reasons. The production of reasons can sometimes benefit from their evaluation "off-line" done to pretest their strength, but the converse doesn't seem to be true; to evaluate reasons, you don't have to imagine yourself producing them. In general, the study of the production of reasons presents a stronger challenge to research on reason than the evaluation of reason (Mercier, 2012).

Reason with limited resources works well in the right social setting

In the interactionist account, reason is not a mechanism or a system superior to intuition. It is a mechanism of intuition about reasons. Like all cognitive mechanisms, its benefits are weighted against its cost. It evolved under a pressure to optimize not its cognitive benefits but its cost-benefit ratio relative to that of other cognitive mechanisms with which it competes for processing resources. Outside of communicative interaction, we claim, the deployment of reason is unlikely to provide an adequate cost-benefit ratio. The social environment provides cues regarding when reason should be triggered, and how to find reasons appropriate in the situation.

Only in the right social setting can one expect reason to function well: people should help each other find progressively better reasons (see Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993), bad reasons should be shot down, and good ones carry the day. What is the right social setting? People who have time to talk with each other, ideally in small groups (Fay, Garrod, & Carletta, 2000), who share some common goals—otherwise, like poker players, they have no incentive to communicate in the first place—people, who share many inferential procedures—otherwise they can't understand each other's reasons—and who disagree on some point—otherwise reasons supporting the consensual view risk piling up unexamined, leading to group polarization or to groupthink (Janis, 1982).

A considerable amount of evidence shows that, in such social setting, reason does indeed function well. In particular, the best ideas present in a group can spread, through discussion, until everyone accepts them. Good insights can even be combined to form a better conclusion than what even the best group members would have been able to reach on their own. These positive outcomes are well-established when small groups discuss accessible logical or mathematical problems, such as the bat and ball (Moshman & Geil, 1998; Laughlin, 2011; Trouche, Sander, & Mercier, 2014; Claidière et al., 2017). They extend well to a variety of other problems: inductive problems, any sort of academic problem faced by students in schools, but also economic predictions, lie detection, medical diagnoses, and more (for reviews, see Mercier & Sperber, 2011; Mercier, 2016b). Even when ascertaining what the best answer is can be difficult, the exchange of reasons seems to point in the right direction. Juries (mock juries at least) make better-informed verdicts, more in line with the opinion of specialists (Hastie, Penrod, & Pennington, 1983). Citizens debating policies in the context of deliberative democracy experiments usually end up more enlightened, with a better grasp of the issues, and more moderate opinions (e.g., Fishkin, 2009; for reviews, see Mercier & Landemore, 2012; Gastil, 2018).

Conclusion: a bounded reason mechanism?

We share with Cosmides and Tooby and Gigerenzer the view that human rationality is bounded through and through: there is no System 2 or other superior mechanism in the human mind that aims at approximating Olympian rationality. We don't believe, however, that an evolutionary approach that recognizes the fully bounded character of human rationality is committed to seeing the faculty of reason hailed by philosopher as a wholly non-existent mechanism, a kind of psychological phlogiston. Where philosophers erred was in overestimating the power of reason and in misrepresenting its function as a prodigious enhancement of individual cognition.

There is, we have argued, a specialized mechanism that produces metarepresentational and metacognitive intuitions about reasons and that, in spite of major differences, is the best true match for reason as classically understood. Hence we call this mechanism "reason." Reasons,

that is, articulated representations of facts together with the conclusion they support, are a very rare and peculiar object in the universe. This makes reason a highly domain-specific cognitive mechanism. Even so, reason indirectly provides a form of virtual domain generality. While the intuitions provided by reason are only about reasons, these reasons themselves can be about anything that humans can think about. In this respect, reason is comparable to linguistic competence: a very specialized competence that, however, makes it possible to produce and understand utterances about anything. This similarity between language and reason is not an accident: reason is mostly deployed by linguistic means and exploits the virtual domain-generality of language itself.

Reasons, we argue, are not tools for individual, solitary thinking; they are tools for social interaction aimed at justifying oneself or at convincing others and at evaluating the justifications and arguments others present to us. In performing these functions, reason is bounded not only by limited internal resources and external opportunities, but also by obstacles to the social flow of information. In situations of cooperative dialogue, reason can help overcome these obstacle and foster convergence. In the case of entrenched antagonisms, on the other hand, reason can foster polarization.

Acknowledgments

Hugo Mercier's work is supported by grants from the Agence Nationale de la Recherche (ANR-10-LABX-0087 to the IEC and ANR-10-IDEX-0001–02 to PSL). 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

Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., … Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction*, 19(1), 1–46.

Bago, B. (2018). *Testing the corrective assumption of dual process theory in reasoning* (PhD thesis). Paris Descartes, Paris.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90–109.

Bago, B., & De Neys, W. (2019). The smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. Thinking & Reasoning 25(3), 257–299.

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(1), 59–78.

Billig, M. (1996). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge: Cambridge University Press.

Boudry, M., Paglieri, F., & Pigliucci, M. (2015). The fake, the flimsy, and the fallacious: Demarcating arguments in real life. *Argumentation*, 29(4), 431–456.

Claidière, N., Trouche, E., & Mercier, H. (2017). Argumentation and the diffusion of counter-intuitive beliefs. *Journal of Experimental Psychology: General*, 146(7), 1052–1066.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7(1), 28–38.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454–459.

Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment*. Hove: Psychology Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, 11(6), 481–486.

Fishkin, J. S. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford: Oxford University Press.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
Gastil, J. (2018). The lessons and limitations of experiments in democratic deliberation. *Annual Review of Law and Social Science*, 14, 271–292.
Gibbard, A. (1990). *Wise choices: Apt feelings*. Cambridge: Cambridge University Press.
Gigerenzer, G. (2007). Fast and frugal heuristics: The tools of bounded rationality. In D. Koehler & N. Harvey (Eds.), *Handbook of judgment and decision making*. Oxford: Blackwell.
Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.
Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.
Janis, I. L. (1982). *Groupthink* (2nd rev. ed). Boston: Houghton Mifflin.
Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720.
Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory and Cognition*, 6, 107–118.
Köymen, B., Rosenbaum, L., & Tomasello, M. (2014). Reasoning during joint decision-making by pre-school peers. *Cognitive Development*, 32, 74–85.
Kruger, J., & Savitsky, K. (2004). The "reign of error" in social psychology: On the real versus imagined consequences of problem-focused research. *Behavioral and Brain Sciences*, 27(3), 349–350.
Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, 118(1), 97.
Kuhn, D. (1991). *The skills of arguments*. Cambridge: Cambridge University Press.
Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, 62(22), 155–178.
Laughlin, P. R. (2011). *Group problem solving*. Princeton, NJ: Princeton University Press.
Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: The MIT Press.
Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford: Oxford University Press.
Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, 22(4), 280–293.
Mercier, H. (2012). Looking for arguments. *Argumentation*, 26(3), 305–324.
Mercier, H. (2016a). Confirmation (or myside) bias. In R. Pohl (Ed.), *Cognitive illusions* (2nd ed., pp. 99–114). London: Psychology Press.
Mercier, H. (2016b). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, 20(9), 689–700.
Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology*, 21(2), 103.
Mercier, H. (2020). *Not Born Yesterday: The Science of Who we Trust and What we Believe*. Princeton, NJ: Princeton University Press.
Mercier, H., Bonnier, P., & Trouche, E. (2016). Why don't people produce better arguments? In L. Macchi, M. Bagassi, & R. Viale (Eds.), *Cognitive unconscious and human rationality* (pp. 205–218). Cambridge, MA: MIT Press.
Mercier, H., Boudry, M., Paglieri, F., & Trouche, E. (2017). Natural-born arguers: Teaching how to make the best of our reasoning abilities. *Educational Psychologist*, 52(1), 1–16.
Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, 33(2), 243–258.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning*, 4(3), 231–248.
Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin and Review*, 11(6), 988–1010.

Perelman, C., & Olbrechts-Tyteca, L. (1958). *The new rhetoric: A treatise on argumentation*. Notre Dame, IN: University of Notre Dame Press.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11(3/4), 347–364.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1–2), 11–36.

Simon, H. A. (1983). *Reason in human affairs*. Stanford, CA: Stanford University Press.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *The Journal of Consumer Research*, 16(2), 158–174.

Smith, S. M., Fabrigar, L. R., & Norris, M. E. (2008). Reflecting on six decades of selective exposure research: Progress, challenges, and opportunities. *Social and Personality Psychology Compass*, 2(1), 464–493.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39–67). Cambridge: Cambridge University Press.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 117–137). Oxford: Oxford University Press.

Sperber, D., & Baumard, N. (2012). Moral reputation: An evolutionary and cognitive perspective. *Mind & Language*, 27(5), 495–518.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language*, 25(4), 359–393.

Sperber, D., & Mercier, H. (2018). Why a modular approach to reason? *Mind & Language*, 131(4), 496–501.

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

Stanovich, K. E. (2004). *The robot's rebellion*. Chicago: Chicago University Press.

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 289–338). New York: Academic Press.

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). New York: Academic Press.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (p. 19). New York: Oxford University Press.

Trouche, E., Johansson, P., Hall, L., & Mercier, H. (2016). The selective laziness of reasoning. *Cognitive Science*, 40(8), 2122–2136.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, 143(5), 1958–1971.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge: Cambridge University Press.