teorema
Vol. XXXVIII/1, 2019, pp. 139-156
ISNN 0210-1602
[BIBLID 0210-1602 (2019) 38:1; pp. 139-156

Replies to Critics

Hugo Mercier and Dan Sperber

We are very thankful to our colleagues who have provided such thoughtful and constructive discussions of our book, *The Enigma of Reason*. Since these commentaries each raise a different set of issues, we respond to them one by one.

KAROLINA PROCHOWNIK raised three questions, each a great invitation for further elaboration. Could reason serve some other social function, she asks, besides argumentation and justification? She suggests two ways in which this could be the case.

Prochownik starts by suggesting that "reason could help people to decide with whom to make alliances and facilitate the recognition of group members. In particular, the evaluation of reasons provided by others to justify their views or actions, in addition to testing their quality and the level of commitment of the reasoner, could serve to filter individuals who really share the audience`s views from those who don't." We agree that reasons can be used to signal one's ideological or, more generally, one's coalitional allegiances. However, we would say that the cognitive mechanisms that influence what contents we express as a function of how these contents inform others of our allegiances are distinct from reason per se. The same mechanisms would be at play whether we use reasons, plain statements of opinions, reactions of approval or disapproval to opinions expressed by others, or any other communicative act (such as wearing a religious or political sign) to serve this end. As a result, the use of reasons to communicate allegiances might not have directly influenced the evolution of our faculty of reason, and thus would not be part of its function in the evolutionary sense of the term, even if it can, on occasions, be a goal or even the main goal, of putting forward reasons.

In the same vein, Prochownik mentions that "when alliances or groups are already formed, public production and sharing of reasons can help to shape group identity (as built upon shared views, behaviors and reasons to hold or perform them), decide common goals, and motivate joint action." Again, we agree with her that such goals can be served by producing reasons. As Prochownik points out, the fact that reason can be productively used to define group identity even by closely knit groups who "share views on moral, political, religious or any other matters" seems to go against our suggestion that such groups are prone to polarization. Actually, there is no contradiction. The exchange of reasons in a like-minded group can bear, for instance, on practical issues that aren't agreed on. In such a case, it should not lead to polarization (for an extended discussion, see Mercier, submitted). One of our favorite examples — continuing on the theme of the abolition of slavery broached in *The Enigma of Reason* — is that of a group of twelve Quakers who would prove formidably influential in defending the abolition of the slave trade in the U.K. As described by Hochschild, these "twelve were of one mind", and, as a result, they could easily have become polarized to the point, say, of defending the immediate abolition of slavery and emancipation of all slaves — a goal unfortunately rather unrealistic at the time. Instead, their meetings led them to focus on more realistic goals, such as the end of the slave trade in the U.K. One possible explanation for the efficiency of argumentation even in conditions that seem propitious for polarization, is that, while the twelve Quakers agreed on a common ultimate goal, they disagreed on the best way to reach it. As a result, argumentation played a useful role in helping them work out the best means to reach the agreed upon end.

Exchange of arguments might also have reminded these twelve Quakers of how right their cause was — even if they all agreed on that — and bolstered their morale. This effect of the exchange of arguments might be closer to what Prochownik has in mind. We would say that, as in the case of reasons used to mark our allegiances, the use of reasons to bolster group morale would not be specific to reasons and would thus not have borne directly on the evolution of the faculty of reason, even if bolstering morale can certainly be the proximal function of some reasons.

Prochownik's second suggestion is that "the evaluation of reasons provided by a speaker [might be] sensitive to the specific features of the social context the reasons are communicated in (e.g., who is reasoning, whom is this person to us, what is the record of our past interactions, what kind of interactions we aim for with that person in the future)."

Our predictions on that point would be as follows. The status of the speaker — are they someone we respect, etc. — should influence how much effort we're willing to put in understanding their reasons. However, for a given amount of effort attributed to processing some reasons, the output of the evaluation process should be similar regardless of the source of the reason, a point classically made by pointing out that a mathematical proof put forward by a scoundrel is just as convincing to a competent audience as it would be if put forward by a paragon of honesty. We have attempted to test this hypothesis by providing people with equally strong arguments coming from sources that were either highly trusted or highly distrusted. This manipulation did not seem to affect participants' evaluation of the arguments [Trouche, Shao, & Mercier (2019)]. We would be tempted to reconcile this theorizing, and these preliminary results, with Prochownik's observation that "people may be more likely to accept or reject other people`s reasons when they have other social reasons to do so" by pointing out that reason's verdict on whether a given reason is good or not is not the only factor that affects our decision to overtly agree or disagree with this reason. An employee might opine to their boss's asinine reason. A political party member might vocally reject arguments challenging their party's platform, even if they are internally bothered by their cogency.

Prochownik's comment also raises an interesting question not about reason itself but about theories of reason. She asks "why the intellectualist theories of reason have been so culturally successful and resistant to being refuted for around 2000 years?" We're obviously hardly the first to stress the importance of the social dimension of reason [see, e.g. Billig (1996); Gibbard (1990); Haidt (2001); Kuhn (1991); Perelman & Olbrechts-Tyteca (1958); Piaget (1928)], but it seems that more individualistic approaches dominate Western psychology and philosophy. If we assume that something along the lines of our interactionist theory is indeed the right take on reason, then one of its predictions should be that "wrong theories … gradually disappear in the process of critical evaluation and discussion by the scientific community," and thus that some version of the interactionist theory should become dominant. Well, we are hopeful that this might come to pass!

We don't have a worked out explanation for the success of intellectualist theories (again, assuming they are essentially mistaken). Still one relevant empirical factor is that humans are good at argumentation but bad at thinking about argumentation in the abstract. People consistently underestimate the efficiency of the exchange of arguments taking place

in group discussions [Mercier, Trouche, Yama, Heintz, & Girotto (2015)]. This is an instance of a more general difficulty humans have with thinking abstractly about the reliability of socially distributed information. They are very good at aggregating information from different social sources — taking into account how many people support an opinion, how competent and benevolent these people are, etc. — but they are terrible at thinking about the aggregation of information from different social sources [Mercier (in press)]. For instance, participants behave nearly optimally when they have to update their answer to a problem as a function of how many other participants have given a different answer [Morgan, Rendell, Ehn, Hoppitt, & Laland (2012); for review, see Mercier & Morin (submitted)]. However, when presented with an abstract voting situation in which we expect the Condorcet Jury Theorem to apply, participants don't believe that the majority of a group is more likely to be right than its average member ([Mercier, Dockendorff, & Schwartzberg, (submitted)]. We shouldn't particularly expect participants to have an abstract understanding of the epistemic benefits of information aggregation mechanisms. However, the fact that they appear to consistently underestimate the efficacy of majority rule, averaging, and argumentation might help explain why theories that stress the importance and the value of the social exchange of ideas have proven less convincing than they arguably should have.

LAURA MACCHI and MARIA BAGASSI [L&M] comment our book from a perspective inspired by the pioneering work of Giuseppe Mosconi (1990), a perspective to the development of which they have contributed in a series of original papers. There are interesting commonalities between their perspective and ours. In particular, we share the view that the role of logic in thinking and reasoning has been overestimated and that the role of language with its conceptual richness and pragmatic resources has been underestimated. Manifestly there are also important differences. We focus our discussion on a couple of clear disagreements, one fairly general, the other about the interpretation of interesting experimental work with the horse-trading task [Maier & Solem (1952)].

M&B object to our claim that in argumentation, logic is mainly used for rhetorical purposes, to streamline and highlight the structure of non-demonstrative arguments, rather than to try and give a genuine logical demonstration of their conclusions. Their objections to our claim, however, are not about the rhetoric of argumentation but about complexity of thought. To the extent that we understand it properly, this ob-

jection is grounded in a genuine disagreement that M&B do not articulate: they see talk of thinking, reasoning, and argumentation as addressing the same range of phenomena from overlapping perspective. We argue at length in our book that there is a distinct, typically human mental mechanism that performs its function in communicative interactions rather than in individual cognition. This mechanism produces reasons for social consumption in the form of justifications and arguments. When we talk of reason, we refer just to this mechanism. We don't see this mechanism as responsible for human inferences in general. Hence, the interesting remarks M&B make about the role of language and context in individual cognition are no objections to our claim.

An original claim we make in our book is that people are able to be much more objective when evaluating arguments presented to them by others than in producing arguments to convince others. One of the predictions that follows from that claim is that, in an exchange of arguments among people with different solutions to some problem but a common interest is solving it, while most individual members of the group are likely to produce poor arguments, these arguments are likely to be recognized as poor by the other members and to be eliminated so that, eventually, the group is likely to converge on the right solution. There are dozens of experiments confirming this prediction. When cooperative groups try to solve demonstrative problems, a single individual with the correct answer is nearly always able to convince the others; in other terms 'truth wins' is by far the most common outcome [Moshman & Geil (1998); Trouche, Sander, & Mercier (2014); for review, see Laughlin (2011)].

M&B, however, offer apparent counter-evidence to our prediction. Using the horse-trading problem, Macchi and Bagassi (2015) showed that participants who had given a wrong answer to a simple accounting problem didn't change their mind when they were presented with the correct solution unless this correct solution was explained in a way that addressed the source of their error. This is interesting in itself but we don't see how it warrants the conclusion that "usually, in fact, the best idea does not prevail, but only the most shared one."

There are two main reasons why the interesting results M&B discuss are no counter-evidence to our prediction:

(1) Our prediction is about group interaction. This experiment does not involve a group discussion but just an interaction between the participant and the experimenter. M&B do report that an earlier study with group discussion [Mosconi, Bagassi, & Serafini

(1988), which alas we could not find] produced similar findings, but then Maier and Solem (1952) report that participants who had given the wrong answer to the same problem changed for the correct one when it was presented in their group. All this is hard to interpret without having greater information on the procedures that were followed.

(2) Most interesting in M&B's study is their demonstration that the mistake of some of the participants was based on taking into account an irrelevant link between two transactions (the protagonist of the story twice buys and then sells a horse at a profit, his total profit being the sum of the two; irrelevantly, it is the same horse that is bought and sold by the protagonist, and it is bought the second time at a price higher than the price at which it has been sold, a fact that participants interpret as a loss that they then wrongly subtract from the overall profit). Participants who made this mistake were not convinced when presented with the correct computation. Note that this correct computation was included in their own computation, to which they wrongly added another computational step. So, it is not surprising that they also needed a non-misleading presentation of the fact that had misled them in order to change their mind. Interesting indeed, but not something on which our approach to reasoning commits us to any prediction.

Although SALVADOR MASCARENHAS is in broad agreement with our thesis, he suggests that, in the intellectualist theory, we might have constructed a bit of a straw man. More specifically, he reminds us that the tradition of the psychology of reasoning, and of judgment and decision making, does not only bear on reason (or System 2, as it might be called), but on inferences more generally — for instance, the first-order inferences that might draw participants towards the 10c answer to the bat and ball, or towards the *p* and *q* cards in the standard Wason selection task. As a result, when these psychologists claim that human reason mostly serves individual, meliorative functions, they would mostly be correct, as most of these first-order inferences do serve such functions.

We acknowledge, with Mascarenhas, that much important work on inferential mechanisms other than reason has, since the 1950s, been done under the heading of 'psychology of reasoning.' This sub-disciplinary framing however may not have been that helpful: the study of specific types of inference was seen as primarily relevant as a source of evidence

for one or another view of reasoning ('mental logic' or 'mental models' in particular) and other questions about the mechanisms involved — their specific procedures, their function, their evolved underpinnings, for instance — got little or no attention.

Moreover, we also believe that distinguishing between reason and other intuitive processes, has been one of the valuable insights stemming from the field (even if we do not share the System 1/System 2 distinction). In this context, our target in *The Enigma of Reason* was the view that some psychologists have formulated of the function of reason in a more restricted sense — what they would call System 2. When Kahneman, Stanovich, or Evans, in some of their writings, suggest that System 2's main function is to correct the mistakes of our first-order inferences, and to help us reach better decisions and better beliefs on our own more generally, their claims clearly do not bear on first-order inferences across the board. As a result, even if our book focuses on reason by contrast with first-order inferences, we do not believe that it represented the field unfairly.

Finally, we want to stress that our thesis regarding the social functions of reason only applies to reason as we define it, and not to other inferential mechanisms. Throughout the book, we insist that reason is only one sub-type of inference, or more precisely a sub-type of metarepresentational inferences, which are themselves a subtype of inference.

CATHAL O'MADAGAIN suggests that we underestimate the importance of cultural transmission in the development of human reason. He points out that some of reason's most striking achievements — such as advanced mathematics — are the outcome of a cultural ratchet effect: a human peculiarity that allows us to add on to the cultural edifices built by our ancestors, potentially reaching new heights at each step. O'Madagain also points out the importance of learning in more mundane forms of reasoning, as when children learn to discount, say, *ad populum* arguments thanks to a parent or teacher pointing out their weakness.

On the whole, these points are well taken. No one could rediscover modern mathematics on their own — it is indeed built on the shoulder of giants. The contrast between the travails that were necessary for advanced mathematics to emerge, and the relative ease with which (some) students can incorporate past discoveries is striking. But what is learnt, exactly, when one learns new mathematical principles? In some cases at least, it's quite plausible that little or no reasoning is involved. Students might apply the rules for long division with very little understanding of why they work.

When it comes to reasons proper, as in a mathematical proof, is it right to say that one has learnt to reason in a different way? Clearly, one has learnt to use specific reasons in a new manner, since this peculiar arrangement of premises to support a given conclusion would likely never have occurred without prompting. Moreover, vaguely echoing Plato, we would argue that the first-order inferences necessary to understand the soundness of the proof must have been present before the proof was encountered. For the proof to be genuinely understood, students must be able to grasp each step. What is new isn't each inferential step, but the fact that these specific inferences are called in, in this specific order, to bear on this specific conclusion. In some cases at least, people are then able to recreate the series of inferences themselves if they have to convince someone else in turn [Claidière, Trouche, & Mercier (2017)].

Still, it is true that we can develop shortcuts, such that a link between a premise and a conclusion that was not intuitive at first—it required going through several intermediate intuitive steps — can become more intuitive with use. It then becomes easier to embed these reasons into increasingly complex chains of reasoning. This is a fascinating phenomenon, and we join O'Madagain in wishing for more research in this direction.

The learning of more mundane arguments is also fascinating — and equally understudied. Again, it is important to try to pinpoint what exactly is being learnt when, say, a child encounters a new argument. O'Madagain takes the example of the *ad populum*, and of its counter-arguments (e.g. if all your friends jumped from a bridge, would you join them?). First, we disagree that the *ad populum*, or other so-called fallacies of argumentation, are by nature fallacious. Following the majority is a sound heuristic in a wide range of cases [Condorcet (1785); Hastie & Kameda (2005)]. As a result, attempting to teach children that *ad populum* arguments are fallacious as a rule might do more harm than good [even if the most likely effect is to do nothing, see Mercier, Boudry, Paglieri, & Trouche (2017)]. Moreover, it seems that participants are well able to discriminate between weaker and stronger forms of these so-called fallacies, even without explicit teaching [Hahn & Oaksford (2007)]. In the case of the *ad populum*, participants (including children) might be able to rely on well calibrated intuitions about when they should follow majority opinions [Morgan, Laland, & Harris (2015); Morgan et al. (2012); for review, see Mercier & Morin (submitted)]. As a result, we suggest that the evaluation of arguments, as a rule, does not require explicit teaching.

The case of argument production is more interesting. Even if children are able to evaluate arguments without explicit learning, they need a

lot of feedback when it comes to argument production. As we suggest in *The Enigma of Reason*, the importance of this feedback has likely influenced how reason works, making us 'lazy' when it comes to producing arguments: if easily available arguments persuade their intended audience, there is no point bothering to find better ones; if they do not, then we can rely on the audience to point that out and, more often than not, to provide a counter-argument that will help us improve our own arguments.

Children, like adults, rely on this 'negative feedback' to learn which arguments do not work and, eventually, stop giving them altogether (as when a child learns that "because I want it" isn't a good argument for grabbing something of off their sibling). But the feedback that O'Madagain is more interested in is positive: how children might appropriate for themselves the arguments used by others. Indeed, when a child (or an adult for that matter) encounters a new argument, they are able not only to evaluate it and decide whether they should change their mind, but also, to some extent, to keep the form of argument in mind and use it in turn [see, e.g. Anderson et al. (2001)]. The inferences children (or, again, adults) have to draw in order to decide the circumstances under which the new argument would be relevant have not, to the best of our knowledge, been studied. Clearly, the learner brings a lot to the table. They do not merely copy the initial argument, as they would if they used it only to defend the specific conclusion the argument supported when they first encountered it. Learners do not either generalize the argument to any potential conclusion (otherwise most arguments would be patently absurd). A process that might look like pure learning — i.e. incorporating into one's argumentative arsenal the arguments one encounters — requires a lot of pre-existing intuitions.

Even if we wholeheartedly agree with O'Madagain's injunction that more attention should be paid to the interactions through which children may learn to reason better, our stance would be that he puts too much weight on learning. In this respect, we should keep in mind that the practice of adults exchanging reasons with children is essentially a peculiarity of middle- and upper-classes in rich societies. In other cultures, adults do not talk with children all that much, when they do, they mostly rely on imperatives, and they certainly do not feel compelled to justify their requests [see, e.g. Gauvain, Munroe, & Beebe (2013); Maratsos (2007); Pye (1986)]. Yet this doesn't seem to stop children growing in traditional cultures from developing the skills to evaluate (at least some) arguments [Castelain, Bernard, Van der Henst, & Mercier (2016)].

These latter results, suggesting that children in traditional cultures differentiate (some) strong from weak arguments, might be partly contested. As O'Madagain points out, experiments showing that young children, including preschoolers and even 2-year-olds [Castelain, Bernard, & Mercier (2018); Koenig (2012); Mercier, Bernard, & Clément (2014); Mercier, Sudo, Castelain, Bernard, & Matsui (2018)] are able to discriminate between strong and weak arguments suffer from a potential confound. Even though the children are presented with arguments, it is possible that they do not understand the statements *qua* arguments, processing instead the premise as if it were a piece of information independent from the conclusion it was offered to defend. This is true, and we are developing ways of testing whether young children can understand arguments *qua* arguments. However, as a rule, in communication at least, production tends to follow evaluation, and children this age do produce arguments [for review, see Mercier (2011), (2016)]. We do not believe therefore that our interpretation in terms of genuine argument evaluation is particularly farfetched.

More generally, while we have argued at length that reason is an evolved adaptation, we keep an open mind about the ways and the degree to which it may be modified by cultural inputs. We conclude our book by stating that "much more must be done to find out to what extent and in which ways [reason] can be harnessed, enriched, and codified differently in various cultural traditions" (p. 334). We welcome O'Magadain's relevant suggestions in this respect.

IRA NOVECK in his friendly, thoughtful, and demanding review makes three reasonable requests: (i) that we defend our deflationary view of logical inferences, (ii) that we develop clear means of falsifying our theory, (iii) that we offer guidelines for how to deal with the justifications participants offer in experiments.

First, let us reassure Noveck that we do not throw the baby of spontaneous deduction with the water of mistaken accounts of the place of deduction in inference generally and reasoning in particular. Not only do we agree with Noveck that spontaneous deduction has been well established and occurs quite commonly, we would even argue that its prevalence is underestimated. Spontaneous deductions studied in the psychology of reasoning are almost exclusively deductions licensed by the presence of a so-called logical term, for instance a connective such as "or" in (1) where (a) entail (b):

(1) (a) Ira is in Paris or in London; he is not in London
(b) Ira is in Paris

We would argue that most items in the lexicon of any ordinary language license some deductions. For instance, in (2), (a) entails (b) in virtue of the semantics of so-called non-logical terms such as the verb "kill" in (4). When such entailments are potentially relevant, they are, we suggest, spontaneously computed:

(2) (a) John killed the dog
(b) The dog is dead

A deduction is a logical derivation from premises to a conclusion that necessarily follows from the premises. It is an abstract relationship, just as is, say, a multiplication. Such abstract relationships can be mentally represented. When we speak of spontaneous deductions, we speak of the process of mentally representing a logical deduction. Mentally represented deductions are not in and of themselves inferences. An inference, as we use the term, is a psychological process that results in the formation (or change in strength) of beliefs or decisions. Inferences may but need not include deductive steps. Making a mental deduction, spontaneously or deliberately, may contribute to the formation of a new belief (or to the confirmation or rejection of a belief already held) but it needs not do so. Even when a deduction plays a role in an inference, the conclusions of the inference may be quite different from those of the deduction. Suppose you are told, "Ira is in Paris or in London; he is not in London." You may spontaneously deduce, "Ira is in Paris," but if you happen to be confident that Ira is not in Paris, what you will infer from this deduction is that what you were told is false. Or suppose that Peter tells you, "John killed the dog," and that (a) you already knew that the dog was dead; (b) you didn't know that Peter knew that the dog was dead; and (c) you don't believe that Peter is in a position to know how the dog died. In that case, the spontaneous deduction from Peter's utterance would indirectly inform you that Peter knew that the dog was dead, but it wouldn't modify your beliefs about what happened to the dog. Generally speaking, the study of mental deductions and that of inference (and in particular or reasoning) should be sharply distinguished. The study of deduction has some relevance to that of inference, but its main relevance may be elsewhere.

Because deductive relationships are an essential aspect of linguistic meaning, deduction — in the sense of the mental representation of these

relationships — often occurs spontaneously and plays, we would argue, an important role in verbal comprehension. We take, however, a strongly pragmatic approach to the role of linguistic meaning in comprehension: linguistic meaning is a piece of evidence from which the audience can infer the speaker's intended meaning; it is not an encoding of that meaning [Wilson & Sperber (2012)]. The main relevance of spontaneous deduction, we suggest, is to the semantics and pragmatics of natural language. In particular, as we argue in the book, a deductive format is often used to highlight the structure of a probabilistic argument. This rhetorical role of deduction in the pragmatics of argumentation is interesting in its own right, but it is, of course, less central and less grand than the role attributed to deduction in logistic approaches to reasoning.

Noveck also mentions interesting recent work [Cesana-Arlotti et al. (2018)] suggesting that pre-verbal infants can detect violation of disjunction elimination. More generally, isn't logical deduction something that humans do not only with natural language but also in non-linguistic thinking (or with 'language of thought')? Don't other animals also use logic? According to Jerry Fodor, they must: "Darwinian selection guarantees that organisms either know the elements of logic or become posthumous" [Fodor (1981), p. 121)]. We argue that inferences that psychologists or philosophers can schematize as deductive need not be so, for at least two reasons: they may be probabilistic rather than deductive, and they may be performed by specialized mechanisms that exploit empirical regularities without representing them as premises in anything resembling a deduction.

So, to answer Noveck's first request in a nutshell: while deductive relationships clearly play a major role in linguistic semantics, the exact role or roles they play in inference in general or in reasoning in particular has been exaggerated and obscured by logicist dogma and is in need of an open-minded reconsideration.

Noveck's second request is that we develop clear means of falsifying the interactionist theory. Our theory rests on some evolutionary, functional hypotheses, namely, that reason evolved by serving justificatory and argumentative functions. Although there are several ways of testing evolutionary hypotheses — for instance looking at variations in fitness, or at genetic data — in the case at hand we have relied on the match between structure and function. Within a broadly adaptationist framework, we expect biological mechanisms to have a structure that fits their function, that serves it well [Williams (1966)]. Our argument thus rests on two sets of claims: (i) claims regarding what the structure of reason is, and (ii) claims regarding the fit between this structure and the

purported functions of reason. Our evolutionary hypotheses could thus be falsified by showing that either of these claims are mistaken.

Regarding the structure of reason, the most relevant traits, for our purposes, were summarized in Table 2, p. 235 of *The Enigma of Reason* (reproduced as Table 1 of the précis). One way of falsifying our hypotheses is thus to show that human reason does not, in fact, possess these traits. For instance, the claim that people are able to evaluate reasons that challenge their point of view objectively is not consensual (to say the least). Evidence that people are deeply biased against such reasons would argue against our hypotheses.

Even if one were to accept our characterization of the traits of reason, one could still attempt if not to falsify, at least to undermine our hypotheses by showing that these traits are better explained as unavoidable by-products of other traits, or that they fit better with another purported function. For example, some psychologists have suggested that the myside bias (or confirmation bias) is a natural outcome of cognitive processing across the board. If that were true, then the existence of this bias could not argue in favor of our hypotheses anymore. We believe this is not true (see Chapter 11), but if it were shown that the myside bias is widespread in our cognitive system, instead of being restricted to reason as we claim, our hypotheses would be weakened. Likewise, people have suggested other functional explanations for the myside bias and; if their explanations were better supported than ours, our hypotheses would again be weakened.

Without falsifying the overall framework, it is possible to falsify more specific claims related to our theory. Noveck mentions the issue of small group discussions: when does these discussions improve performance on reasoning and decision making problems? Here, our theory makes broad predictions: on the whole, when groups of people who disagree on a given point, but share some common incentives, exchange arguments together, the best arguments should carry the day and performance should improve. However, there will be exceptions. For example, sometimes the participants who have the correct answer might have reached it partly by chance, or without being able to formulate reasons defending their correct answer in a way that would convince other participants who came to a wrong solution in a roundabout way (as might be the case for the horse seller problem presented by Macchi & Bagassi). Such special instances do not invalidate our theory, since the theory doesn't claim that people are always able to turn their correct intuitions into reasons apt at convincing the intended audience. Still, the theory requires some ancil-

lary considerations to explain why argumentation fails to work in these situations (i.e. because people don't seem to be able to articulate why the correct answer is correct) and, if such situations were more common than situations in which argumentation works well, our theory would be in trouble.

Another example is that of 'reason-based choice.' Relying on a long tradition in judgment and decision making, we claim that many deviations from normative behavior in decision making tasks reflect the operation of reason, as it drives participants towards the most easily justifiable answer — whether it is otherwise the best answer or not (see Chapter 14). In each specific case, this explanation can be tested, for instance by looking at whether (i) making participants accountable amplifies the deviation from normative behavior (it should), (ii) cognitive load reduces the deviation from normative behavior (it should), (iii) non-human animals make the same mistakes (they shouldn't), or (iv) the justifications participants produce are in line with the answers they give, and appear somewhat convincing to others (they should).

This last point brings us to Noveck's final request regarding the usefulness of justifications in psychological research. First, a bit of terminology: we would differentiate explanations from justifications. Justifications are based on reasons, whereas explanations may use reasons, but they need not. For instance, if a participant explains their answer in a psychological experiment by saying 'I'm too tired to think', this is an explanation, but not a justification (not a reason). We'll take it that when Noveck talks about explanations, what he has in mind is justifications — which would indeed form the majority of participants' answers, whether they are asked to explain or to justify their answers. (Given that one of us has used 'explain' in prompts, even though he was looking for justifications, we can see how this might not have been clear!)

With this in mind, the answer to the question "Does it follow that a justification does not count as reasoning when the task simply asks a participant to provide one, as they are in the first phase of the Trouche et al. (2015) experiment?" is a qualified no. We spend some time arguing that the same reasons can be used either or both as justifications or as arguments, with no clear demarcation between these two uses. So, justifications are indeed reasons, produced by our faculty of reason. Participants who give a reason to justify their answer to the experimenter could as well use it to convince another participant to agree with them. The fact that these reasons likely played no causal role in arriving at the answers is normal: we claim that this is generally true of the reasons we produce.

By contrast, the question of whether the cues described by Noveck ("Please note that the sum of the two percentages must be 100%. For example, Anne is either a mathematics teacher or a French literature teacher" to help participants solve a probabilistic problem) are reasons is more problematic. First, there is no intrinsic quality in a statement that makes it a reason or not, it all depends on how it is processed — so that the same statement can be processed as an implicit instruction to follow in producing an answer by a participant and as a reason to be invoked in justifying one's answer by another participant (in the same way that an imperative utterance can be processed as an order by someone and as an advice by someone else, say — this ambiguity raises difficulties in interpreting some experimental results, see the discussion of O'Madagain's final point above). Second, in the case at hand, there are no strong elements helping participants process the statement as a reason: no explicit conclusion, no connectives, etc. As a result, it's quite plausible that participants would interpret the cue as a piece of information, from which they are free to draw whatever conclusion they want. If that is the case, a prediction would be that if the conclusion were spelled out, and the link between the cue (which becomes a premise) and the conclusion clarified, more people should be able understand the relevance of the cue, and thus to reach the correct answer.

Unfortunately, testing whether or not a given statement is processed as a reason is not trivial. As mentioned above, we're working on experimental paradigms that would allow such testing in children. Other methods, aimed primarily at adults, using reaction times, eye tracking, or even neuroscientific tools might be used as well, but they still have to be developed (another work in progress!).

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

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), pp. 1-46.

BILLIG, M. (1996), *Arguing and Thinking: A Rhetorical Approach to Social Psychology*; Cambridge: Cambridge University Press.

CASTELAIN, T., BERNARD, S., & MERCIER, H. (2018), "Evidence that Two-Year-Old Children are Sensitive to Information Presented in Arguments"; *Infancy*, *23*(1), 124–135.

CASTELAIN, T., BERNARD, S., VAN DER HENST, J.-B., & MERCIER, H. (2016), "The Influence of Power and Reason on Young Maya Children's Endorsement of Testimony; *Developmental Science*, *19*(6), pp. 957-966.

CESANA-ARLOTTI, N., MARTÍN, A., TÉGLÁS, E., VOROBYOVA, L., CETNARSKI, R., & BONATTI, L. L. (2018), "Precursors of Logical Reasoning in Preverbal Human Infants"; *Science*, *359*(6381), pp. 1263-1266.

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

CONDORCET. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'imprimerie royale.

FODOR, J.A., (1981), *Representations: Philosophical Essays on the Foundations of Cognitive Science*; Cambridge: MIT Press.

GAUVAIN, M., MUNROE, R. L., & BEEBE, H. (2013), "Children's Questions in Cross-Cultural Perspective a Four-Culture Study"; *Journal of Cross-Cultural Psychology*, *44*(7), pp. 1148-1165.

GIBBARD, A. (1990), *Wise choices, Apt feelings*; Cambridge: Cambridge University Press.

HAHN, U., & OAKSFORD, M. (2007), "The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies"; *Psychological Review*, *114*(3), pp. 704-732.

HAIDT, J. (2001), "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment"; *Psychological Review*, *108*(4), pp. 814-834.

HASTIE, R., & KAMEDA, T. (2005), "The Robust Beauty of Majority Rules in Group decisions"; *Psychological Review*, *112*(2), pp. 494-508.

HOCHSCHILD, A. (2006), *Bury the Chains: Prophets and Rebels in the Fight to Free an Empire's Slaves*. Boston: Houghton Mifflin Company.

KOENIG, M. A. (2012), "Beyond Semantic Accuracy: Preschoolers Evaluate a Speaker's Reasons"; *Child Development*, *83*(3), pp. 1051-1063.

KUHN, D. (1991), *The Skills of Arguments*; Cambridge: Cambridge University Press.

LAUGHLIN, P. R. (2011), *Group Problem Solving*; Princeton: Princeton University Press.

MACCHI, L., & BAGASSI, M. (2015), "When Analytic Thought is Challenged by a Misunderstanding; *Thinking & Reasoning*, *21*(1), pp. 147-164.

MAIER, N. R., & SOLEM, A. R. (1952), "The Contribution of a Discussion Leader to the Quality of Group Thinking: the Efective Use of Minority Opinions"; *Human Relations*, 5(3), pp. 277-288.

MARATSOS, M. P. (2007), "Commentary"; *Monographs of the Society for Research in Child Development*, *72*, pp. 121-126.

MERCIER, H. (in press), "A paradox of Information Aggregation: We Do It Well but Think About It Poorly, and Why this is a Problem for Institutions"; In N. Ballantyne & D. Dunning (Eds.), *Epistemology and Psychology*. New York: Oxford University Press.

— (submitted), *Not Born Yesterday: The Science of Who We Trust and What We Believe*; New York, Princeton University Press.

— (2011) "Reasoning Serves Argumentation in Children"; *Cognitive Development*, *26*(3), pp. 177-191.

— (2012), "Looking for Arguments"; *Argumentation*, *26*(3), pp. 305-324.

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

MERCIER, H., BERNARD, S., & CLÉMENT, F. (2014), "Early Sensitivity to Arguments: How Preschoolers Weight Circular Arguments"; *Journal of Experimental Child Psychology*, *125*, pp. 102-109.

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), pp. 1-16.

MERCIER, H., DOCKENDORFF, M., & SCHWARTZBERG, M. (submitted), "Democratic Legitimacy and Attitudes About Information-Aggregation Procedures".

MERCIER, H., & MORIN, O. (submitted), "Majority Rules: How Good Are We at Aggregating Convergent Opinions"?

MERCIER, H., SUDO, M., CASTELAIN, T., BERNARD, S., & MATSUI, T. (2018), "Japanese Preschoolers' Evaluation of Circular and Non-Circular Arguments"; *European Journal of Developmental Psychology*, *15*(5), pp. 493-505.

MERCIER, H., TROUCHE, E., YAMA, H., HEINTZ, C., & GIROTTO, V. (2015), "Experts and Laymen Grossly Underestimate the Benefits of Argumentation for Reasoning"; *Thinking & Reasoning*, *21*(3), pp. 341-355.

MORGAN, T. J. H., LALAND, K. N., & HARRIS, P. L. (2015), "The Development of Adaptive Conformity in Young Children: Effects of Uncertainty and Consensus"; *Developmental Science*, *18*(4), pp. 511-524.

MORGAN, T. J. H., RENDELL, L. E., EHN, M., HOPPITT, W., & LALAND, K. N. (2012), "The Evolutionary Basis of Human Social Learning"; *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1729), pp. 653-662.

MOSCONI, G., BAGASSI, M., & SERAFINI, M. G. (1988), "Solutori e benrispondenti. II. Il problema della compravendita del cavallo: Discussione e ricerca di gruppo"; *Giornale Italiano Di Psicologia, XV*, *4*, pp. 671-694.

MOSHMAN, D., & GEIL, M. (1998), "Collaborative Reasoning: Evidence for CollectiveRrationality"; *Thinking and Reasoning*, *4*(3), pp. 231-248.

PERELMAN, C., & OLBRECHTS-TYTECA, L. (1958), *The New Rhetoric: A Treatise on Argumentation*. Notre Dame, IN: University of Notre Dame Press.

PIAGET, J. (1928), *Judgment and Reasoning in the Child*; London: Routledge and Kegan Paul.

PYE, C. (1986), "Quiché Mayan Speech to Children"; *Journal of Child Language*, *13*(1), pp. 85-100.

SPERBER, D. ET AL., (2010), "Epistemic Vigilance"; *Mind and Language*, 25(4), pp. 359-393.

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), pp. 1958-971.

TROUCHE, E., SHAO, J., & MERCIER, H. (2019), "How is Argument Evaluation Biased?"; *Argumentation*, (in press).

WILLIAMS, G. C. (1966), *Adaptation and Natural Selection*; Princeton: Princeton University Press.

WILSON, D. & SPERBER, D (2012), *Meaning and Relevance*; Cambridge: Cambridge University Press.