



**HAL**  
open science

## Why do humans reason? Arguments for an argumentative theory.

Hugo Mercier, Dan Sperber

► **To cite this version:**

Hugo Mercier, Dan Sperber. Why do humans reason? Arguments for an argumentative theory.. Behavioral and Brain Sciences, 2011, 34 (2), pp.57-74; discussion 74-111. 10.1017/S0140525X10000968 . hal-00904097

**HAL Id: hal-00904097**

**<https://hal.science/hal-00904097>**

Submitted on 21 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Why do humans reason? Arguments for an argumentative theory

**Hugo Mercier**

Philosophy, Politics and Economics Program, University of Pennsylvania,  
Philadelphia, PA 19104

hmercier@sas.upenn.edu <http://sites.google.com/site/hugomercier/>

**Dan Sperber**

Jean Nicod Institute (EHESS-ENS-CNRS), 75005 Paris, France; Department  
of Philosophy, Central European University, Budapest, Hungary

dan@sperber.fr <http://www.dan.sperber.fr>

**Abstract:** Reasoning is generally seen as a means to improve knowledge and make better decisions. However, much evidence shows that reasoning often leads to epistemic distortions and poor decisions. This suggests that the function of reasoning should be rethought. Our hypothesis is that the function of reasoning is argumentative. It is to devise and evaluate arguments intended to persuade. Reasoning so conceived is adaptive given the exceptional dependence of humans on communication and their vulnerability to misinformation. A wide range of evidence in the psychology of reasoning and decision making can be reinterpreted and better explained in the light of this hypothesis. Poor performance in standard reasoning tasks is explained by the lack of argumentative context. When the same problems are placed in a proper argumentative setting, people turn out to be skilled arguers. Skilled arguers, however, are not after the truth but after arguments supporting their views. This explains the notorious confirmation bias. This bias is apparent not only when people are actually arguing, but also when they are reasoning proactively from the perspective of having to defend their opinions. Reasoning so motivated can distort evaluations and attitudes and allow erroneous beliefs to persist. Proactively used reasoning also favors decisions that are easy to justify but not necessarily better. In all these instances traditionally described as failures or flaws, reasoning does exactly what can be expected of an argumentative device: Look for arguments that support a given conclusion, and, *ceteris paribus*, favor conclusions for which arguments can be found.

**Keywords:** argumentation; confirmation bias; decision making; dual process theory; evolutionary psychology; motivated reasoning; reason-based choice; reasoning

*Inference* (as the term is most commonly understood in psychology) is the production of new mental representations on the basis of previously held representations. Examples of inferences are the production of new beliefs on the basis of previous beliefs, the production of expectations on the basis of perception, or the production of plans on the basis of preferences and beliefs. So understood, inference need not be deliberate or conscious. It is at work not only in conceptual thinking but also in perception and in motor control (Kersten et al. 2004; Wolpert & Kawato 1998). It is a basic ingredient of any cognitive system. *Reasoning*, as commonly understood, refers to a very special form of inference at the conceptual level, where not only is a new mental representation (or *conclusion*) consciously produced, but the previously held representations (or *premises*) that warrant it are also consciously entertained. The premises are seen as providing *reasons* to accept the conclusion. Most work in the psychology of reasoning is about reasoning so understood. Such reasoning is typically human. There is no evidence that it occurs in nonhuman animals or in preverbal children.<sup>1</sup>

How do humans reason? Why do they reason? These two questions are mutually relevant, since the mechanisms for reasoning should be adjusted to its function. While the how-question has been systematically investigated (e.g.,

Evans et al. 1993; Johnson-Laird 2006; Oaksford & Chater 2007; Rips 1994), there is very little discussion of the why-question. How come? It may be that the function of reasoning is considered too obvious to deserve much

HUGO MERCIER is a postdoctoral fellow at the University of Pennsylvania. His work has focused on the theme of the present article – reasoning and argumentation. He is working on a series of articles that cover this issue from different perspectives – developmental, cross-cultural, political, and historical.

DAN SPERBER is a French social and cognitive scientist. He is professor of philosophy and cognitive science at the Central European University, Budapest, and directeur de recherche emeritus at the Institut Jean Nicod, (CNRS, ENS, and EHESS, Paris). He is the author of *Rethinking Symbolism* (1975), *On Anthropological Knowledge* (1985), and *Explaining Culture* (1996); the co-author with Deirdre Wilson of *Relevance: Communication and Cognition* (1986 – Second Revised Edition, 1995); the editor of *Metarepresentations: A Multidisciplinary Perspective* (2000); the co-editor with David Premack and Ann James Premack of *Causal Cognition: A Multidisciplinary Debate* (1995), and, with Ira Noveck, of *Experimental Pragmatics* (2004).

attention. According to a long philosophical tradition, reasoning is what enables the human mind to go beyond mere perception, habit, and instinct. In the first, theoretical section of this article we sketch a tentative answer to the how-question and then focus on the why-question: We outline an approach to reasoning based on the idea that the primary function for which it evolved is the production and evaluation of arguments in communication. In sections 2–5, we consider some of the main themes and findings in the experimental literature on reasoning and show how our approach helps make better sense of much of the experimental evidence and hence gains empirical support from it.

## 1. Reasoning: Mechanism and function

### 1.1. Intuitive inference and argument

Since the 1960s, much work in the psychology of reasoning has suggested that, in fact, humans reason rather poorly, failing at simple logical tasks (Evans 2002), committing egregious mistakes in probabilistic reasoning (Kahneman & Tversky 1972; Tversky & Kahneman 1983), and being subject to sundry irrational biases in decision making (Kahneman et al. 1982). This work has led to a rethinking of the mechanisms for reasoning, but not – or at least, not to the same degree – of its assumed function of enhancing human cognition and decision making. The most important development has been the emergence of dual-process models that distinguish between intuitions and reasoning (or system 1 and system 2 reasoning) (Evans 2007; Johnson-Laird 2006; Kahneman 2003; Kahneman & Frederick 2002; 2005; Sloman 1996; Stanovich 2004). Here we outline our own dual-process approach: We contend in particular that the arguments used in reasoning are the output of a mechanism of intuitive inference (Mercier & Sperber 2009; Sperber 1997; 2001).

A *process of inference* is a process, the representational output of which necessarily or probabilistically follows from its representational input. The function of an inferential process is to augment and correct the information available to cognitive system. An evolutionary approach suggests that inferential processes, rather than being based on a single inferential mechanism or constituting a single integrated system, are much more likely to be performed by a variety of domain-specific mechanisms, each attuned to the specific demands and affordances of its domain (e.g., see Barkow et al. 1992). The inferential processes carried out by these mechanisms are unconscious: They are not mental acts that individuals decide to perform, but processes that take place inside their brain, at a “sub-personal” level (in the sense of Dennett 1969). People may be aware of having reached a certain conclusion – be aware, that is, of the output of an inferential process – but we claim that they are never aware of the process itself. All inferences carried out by inferential mechanisms are in this sense *intuitive*. They generate *intuitive beliefs*; that is, beliefs held without awareness of reasons to hold them.

The claim that all inferential processes carried out by specialized inferential mechanisms are unconscious and result in intuitive inferences may seem to contradict the common experience of forming a belief because one has reflected on reasons to accept it – and not, or not only,

because of its intuitive force. Such beliefs, held with awareness of one’s reasons to hold them, are better described not as intuitive but as *reflective beliefs* (Sperber 1997). Our consciously held reason for accepting a reflective belief may be trust in its source (the professor, the doctor, the priest). Our reasons may also have to do with the content of the belief: We realize, for example, that it would be inconsistent on our part to hold to our previous beliefs and not accept some given new claim. Far from denying that we may arrive at a belief through reflecting on our reasons to accept it, we see this as *reasoning proper*, the main topic of this article. What characterizes reasoning proper is indeed the awareness not just of a conclusion but of an argument that justifies accepting that conclusion. We suggest, however, that arguments exploited in reasoning are the output of an intuitive inferential mechanism. Like all other inferential mechanisms, its processes are unconscious (as also argued by Johnson-Laird 2006, p. 53; and Jackendoff 1996) and its conclusions are intuitive. However, *these intuitive conclusions are about arguments*; that is, about representations of relationships between premises and conclusions.

The intuitive inferences made by humans are not only about ordinary objects and events in the world. They can also be about representations of such objects or events (or even about higher-order representations of representations). The capacity to represent representations, and to draw inferences about them, is a *metarepresentational* capacity with formal properties relevant to the mental computations involved (Recanati 2000; Sperber 2000b). Several mental mechanisms use this metarepresentational capacity. In particular, humans have a mechanism for representing mental representations and for drawing intuitive inferences about them. This Theory of Mind mechanism is essential to our understanding of others and of ourselves (Leslie 1987; Premack & Woodruff 1978). Humans also have a mechanism for representing verbal representations and for drawing intuitive inferences about them. This *pragmatic* mechanism is essential to our understanding of communicated meaning in context (Grice 1975; Sperber & Wilson 2002).

We want to argue that there is yet another intuitive metarepresentational mechanism, a mechanism for representing possible reasons to accept a conclusion – that is, for representing arguments – and for evaluating their strength. Arguments should be sharply distinguished from inferences. An inference is a process the output of which is a representation. An argument is a complex representation. Both an inference and an argument have what can be called a conclusion, but in the case of an inference, the conclusion is the output of the inference; in the case of an argument, the conclusion is a part – typically the last part – of the representation. The output of an inference can be called a “conclusion” because what characterizes an inferential process is that its output is justified by its input; the way however in which the input justifies the output is not represented in the output of an intuitive inference. What makes the conclusion of an argument a “conclusion” (rather than simply a proposition) is that the reasons for drawing this conclusion on the basis of the premises are (at least partially) spelled out. As Gilbert Harman (1986) has justly argued, it is a common but costly mistake to confuse the causally and temporally related steps of an inference with the logically related

steps of an argument. The causal steps of an inference need not recapitulate the logical step of any argument for it to be an inference, and the logical step of an argument need not be followed in any inference for it to be an argument.

Descartes's famous Cogito argument, "I think therefore I am," illustrates the manner in which an argument can be the output of an intuitive inference. Most people believe intuitively that they exist and are not looking for reason to justify this belief. But should you look for such reasons – that is, should you take a reflective stance towards the proposition that you exist – Descartes's argument would probably convince you: It is intuitively evident that the fact that you are thinking is a good enough reason to accept that you exist, or, in other terms, that it would be inconsistent to assert "I think" and to deny "I am." What is not at all obvious in this particular case are the reasons for accepting that this intuitively good argument is truly a good argument, and philosophers have been hotly debating the issue (e.g., Katz 1986).

As simple as the Cogito or more complex, all arguments must ultimately be grounded in *intuitive* judgments that given conclusions follow from given premises. In other words, we are suggesting that arguments are not the output of a system 2 mechanism for explicit reasoning, that would be standing apart from, and in symmetrical contrast to, a system 1 mechanism for intuitive inference. Rather, arguments are the output of one mechanism of intuitive inference among many that delivers intuitions about premise-conclusion relationships. Intuitions about arguments have an evaluative component: Some arguments are seen as strong, others as weak. Moreover, there may be competing arguments for opposite conclusions and we may intuitively prefer one to another. These evaluation and preferences are ultimately grounded in intuition.

If we accept a conclusion because of an argument in its favor that is intuitively strong enough, this acceptance is an epistemic decision that we take at a personal level. If we construct a complex argument by linking argumentative steps, each of which we see as having sufficient intuitive strength, this is a personal-level mental action. If we verbally produce the argument so that others will see its intuitive force and will accept its conclusion, it is a public action that we consciously undertake. The mental action of working out a convincing argument, the public action of verbally producing this argument so that others will be convinced by it, and the mental action of evaluating and accepting the conclusion of an argument produced by others correspond to what is commonly and traditionally meant by *reasoning* (a term that can refer to either a mental or a verbal activity).

Why should the reflective exploitation of one mechanism for intuitive inference among many stand out as so important that it has been seen as what distinguishes humans from beasts? Why, in dual-process theories of reasoning, should it be contrasted on its own with all the mechanisms for intuitive inference taken together? We see three complementary explanations for the saliency of reasoning. First, when we reason, we know that we are reasoning, whereas the very existence of intuitive inference was seen as controversial in philosophy before its discovery in cognitive science. Second, while an inferential mechanism that delivers intuitions about arguments is,

strictly speaking, highly domain specific, the arguments that it delivers intuitions about can be representations of anything at all. Thus, when we reason on the basis of these intuitions, we may come to conclusions in all theoretical and practical domains. In other words, even though inferences about arguments are domain specific (as evolutionary psychologists would expect), they have domain general consequences and provide a kind of virtual domain generality (without which traditional and dual-process approaches to reasoning would make little sense). Third, as we will now argue, the very function of reasoning puts it on display in human communication.

## 1.2. The function of reasoning

We use *function* here in its biological sense (see Allen et al. 1998). Put simply, a function of a trait is an effect of that trait that causally explains its having evolved and persisted in a population: Thanks to this effect, the trait has been contributing to the fitness of organisms endowed with it. In principle, several effects of a trait may contribute to fitness, and hence a trait may have more than a single function. Even then, it may be possible to rank the importance of different functions, and in particular to identify a function for which the trait is best adapted as its main function. For instance, human feet have the functions of allowing us both to run and to walk, but their plantigrade posture is better adapted for walking than for running, and this is strong evidence that walking is their main function (Cunningham et al. 2010). In the same vein, we are not arguing against the view that our reasoning ability may have various advantageous effects, each of which may have contributed to its selection as an important capacity of the human mind. We do argue, however, that reasoning is best adapted for its role in argumentation, which should therefore be seen as its main function.

There have been a few tentative attempts in dual-process approaches to explain the function and evolution of reasoning. The majority view seems to be that the main function of reasoning is to enhance individual cognition. This is expressed, for instance, by Kahneman (2003, p. 699), Gilbert (2002), Evans and Over (1996, p. 154), Stanovich (2004, p. 64), and Sloman (1996, p. 18). This classical view of reasoning – it goes back to Descartes and to ancient Greek philosophers – faces several problems that become apparent when its functional claims are laid out in slightly greater detail. It is sometimes claimed (e.g., by Kahneman 2003) that the meliorative function of system 2 reasoning is achieved by correcting mistakes in system 1 intuitions. However, reasoning itself is a potential source of new mistakes. Moreover, there is considerable evidence that when reasoning is applied to the conclusions of intuitive inference, it tends to rationalize them rather than to correct them (e.g., Evans & Wason 1976).

According to another hypothesis, conscious reasoning "gives us the possibility to deal with novelty and to anticipate the future" (Evans & Over 1996, p. 154). But giving an organism the possibility to deal with novelty and to anticipate the future is less a characterization of reasoning than it is of learning (or even, it could be argued, of cognition in general). After all, learning can be defined as "the process by which we become able to use past and current events to predict what the future holds" (Niv &

Schoenbaum 2008, p. 265). The issue is not whether, on occasion, reasoning can help correct intuitive mistakes or better adapt us to novel circumstances. No doubt, it can. The issue is how far these occasional benefits explain the costs incurred, and hence the very existence of reasoning among humans, and also explain its characteristic features. In any case, evolutionary hypotheses are of little help unless precise enough to yield testable predictions and explanations. To establish that reasoning has a given function, we should be able at least to identify signature effects of that function in the very way reasoning works.

Here we want to explore the idea that the emergence of reasoning is best understood within the framework of the evolution of human communication. Reasoning enables people to exchange arguments that, on the whole, make communication more reliable and hence more advantageous. The main function of reasoning, we claim, is *argumentative* (Sperber 2000a; 2001; see also Billig 1996; Dessalles 2007; Kuhn 1992; Perelman & Olbrechts-Tyteca 1969; for a very similar take on the special case of moral reasoning, see Gibbard 1990 and Haidt 2001).

For communication to be stable, it has to benefit both senders and receivers; otherwise they would stop sending or stop receiving, putting an end to communication itself (Dawkins & Krebs 1978; Krebs & Dawkins 1984). But stability is often threatened by dishonest senders who may gain by manipulating receivers and inflicting too high of a cost on them. Is there a way to ensure that communication is honest? Some signals are reliable indicators of their own honesty. Costly signals such as a deer antlers or a peacock tail both signal and show evidence that the individual is strong enough to pay that cost (Zahavi & Zahavi 1997). Saying “I am not mute” is proof that the speaker is indeed not mute. However, for most of the rich and varied informational contents that humans communicate among themselves, there are no available signals that would be proof of their own honesty. To avoid being victims of misinformation, receivers must therefore exercise some degree of what may be called *epistemic vigilance* (Sperber et al. 2010). The task of epistemic vigilance is to evaluate communicator and the content of their messages in order to filter communicated information.

Several psychological mechanisms may contribute to epistemic vigilance. The two most important of these mechanisms are trust calibration and coherence checking. People routinely calibrate the trust they grant different speakers on the basis of their competence and benevolence (Petty & Wegener 1998). Rudiments of trust calibration based on competence have been demonstrated in 3-year-old children (for reviews, see Clément 2010; Harris 2007). The ability to distrust malevolent informants has been shown to develop in stages between the ages of 3 and 6 (Mascaro & Sperber 2009).

The interpretation of communicated information involves activating a context of previously held beliefs and trying to integrate the new with old information. This process may bring to the fore incoherencies between old and newly communicated information. Some initial coherence checking thus occurs in the process of comprehension. When it uncovers some incoherence, an epistemically vigilant addressee must choose between two alternatives. The simplest is to reject communicated information, thus avoiding any risk of being

misled. This may, however, deprive the addressee of valuable information and of the opportunity to correct or update earlier beliefs. The second, more elaborate, alternative consists in associating coherence checking and trust calibration and allowing for a finer-grained process of belief revision. In particular, if a highly trusted individual tells us something that is incoherent with our previous beliefs, some revision is unavoidable: We must revise either our confidence of the source or our previous beliefs. We are likely to choose the revision that reestablishes coherence at the lesser cost, and this will often consist in accepting the information communicated and revising our beliefs.

What are the options of a communicator wanting to communicate a piece of information that the addressee is unlikely to accept on trust? One option may be for the communicator to provide evidence of her reliability in the matter at hand (for instance, if the information is about health issues, she might inform the addressee that she is a doctor). But what if the communicator is not in a position to boost her own authority? Another option is to try to convince her addressee by offering premises the addressee already believes or is willing to accept on trust, and showing that, once these premises are accepted, it would be less coherent to reject the conclusion than to accept it. This option consists in producing arguments for one's claims and in encouraging the addressee to examine, evaluate, and accept these arguments. Producing and evaluating arguments is, of course, a use of reasoning.

Reasoning contributes to the effectiveness and reliability of communication by allowing communicators to argue for their claim and by allowing addressees to assess these arguments. It thus increases both in quantity and in epistemic quality the information humans are able to share. Claiming as we do that this role of reasoning in social interaction is its main function fits well with much current work stressing the role of sociality in the unique cognitive capacities of humans (Byrne & Whiten 1988; Dunbar 1996; Dunbar & Shultz 2003; Hrdy 2009; Humphrey 1976; Tomasello et al. 2005; Whiten & Byrne 1997). In particular, the evolutionary role of small group cooperation has recently been emphasized (Dubreuil 2010; Sterelny, in press). Communication plays an obvious role in human cooperation both in the setting of common goals and in the allocation of duties and rights. Argumentation is uniquely effective in overcoming disagreements that are likely to occur, in particular in relatively equalitarian groups. While there can hardly be any archaeological evidence for the claim that argumentation already played an important role in early human groups, we note that anthropologists have repeatedly observed people arguing in small-scale traditional societies (Boehm et al. 1996; Brown 1991; Mercier, in press a).

The main function of reasoning is argumentative: Reasoning has evolved and persisted mainly because it makes human communication more effective and advantageous. As most evolutionary hypotheses, this claim runs the risk of being perceived as another “just so story.” It is therefore crucial to show that it entails falsifiable predictions. If the main function of reasoning is indeed argumentative, then it should exhibit as signature effects strengths and weaknesses related to the relative importance of this function compared to other potential functions of reasoning. This should be testable through

experimental work done here and now. Our goal now is to spell out and explain what signature effects we predict, to evaluate these predictions in light of the available evidence, and to see whether they help make better sense of a number of well-known puzzles in the psychology of reasoning and decision making. Should one fail, on the other hand, to find such signature of the hypothesized argumentative function of reasoning, and even more should one find that the main features of reasoning match some other function, then our hypothesis should be considered falsified.<sup>2</sup>

Several predictions can be derived from the argumentative theory of reasoning. The first and most straightforward is that reasoning should do well what it evolved to do; that is, produce and evaluate arguments (sects. 2.1 and 2.2). In general, adaptations work best when they are used to perform the task they evolved to perform. Accordingly, reasoning should produce its best results when used in argumentative contexts, most notably in group discussions (sect. 2.3). When we want to convince an interlocutor with a different viewpoint, we should be looking for arguments in favor of our viewpoint rather than in favor of hers. Therefore, the next prediction is that reasoning used to produce argument should exhibit a strong confirmation bias (sect. 3). A further related prediction is that, when people reason on their own about one of their opinions, they are likely to do so proactively, that is, anticipating a dialogic context, and mostly to find arguments that support their opinion. Evidence of the existence of such *motivated reasoning* is reviewed in section 4. Finally, we want to explore the possibility that, even in decision making, the main function of reasoning is to produce arguments to convince others rather than to find the best decision. Thus, we predict that reasoning will drive people towards decisions for which they can argue – decisions that they can justify – even if these decisions are not optimal (sect. 5).

## 2. Argumentative skills

### 2.1. Understanding and evaluating arguments

In this section, we review evidence showing that people are skilled arguers, using reasoning both to evaluate and to produce arguments in argumentative contexts. This, in itself, is compatible with other accounts of the main function of reasoning. However, this evidence is relevant because the idea that people are not very skilled arguers is relatively common; if it were true, then the argumentative theory would be a nonstarter. It is therefore crucial to demonstrate that this is not the case and that people have good argumentative skills, starting with the ability to understand and evaluate arguments.

The understanding of arguments has been studied in two main fields of psychology: persuasion and attitude change, on the one hand, and reasoning, on the other. The aims, methods, and results are different in the two fields. Within social psychology, the study of persuasion and attitude change has looked at the effects of arguments on attitudes. In a typical experiment, participants hear or read an argument (a “persuasive message”), and the evolution of their attitude on the relevant topic is measured. For instance, in a classic study by Petty and Cacioppo (1979), participants were presented with arguments

supporting the introduction of a comprehensive senior exam. Some participants heard strong arguments (such as data showing that “graduate and professional schools show a preference for undergraduates who have passed a comprehensive exam”), while others heard much weaker arguments (such as a quote from a graduate student saying that “since they have to take comprehensives, undergraduates should take them also”). In this experiment, it was shown that participants who would be directly affected by the setting up of a comprehensive exam were much more influenced by strong arguments than by weak ones. This experiment illustrates the more general finding stemming from this literature that, *when they are motivated, participants are able to use reasoning to evaluate arguments accurately* (for a review, see Petty & Wegener 1998).

The demonstration that people are skilled at assessing arguments seems to stand in sharp contrast with findings from the psychology of reasoning. In a typical reasoning experiment, participants are presented with premises and asked either to produce or to evaluate a conclusion that should follow logically. Thus, they may have to determine what, if anything, follows from premises such as “If there is a vowel on the card, then there is an even number on the card. There is not an even number on the card.” In such tasks, Evans (2002) recognizes that “logical performance . . . is generally quite poor” (p. 981). To give just one example, it was found in a review that an average of 40% of participants fail to draw the simple *modus tollens* conclusion that was used as an example (if  $p$  then  $q$ , not  $q$ , therefore not  $p$ ) (Evans et al. 1993). However, reasoning, according to the present view, should mostly provide a felicitous evaluation in dialogic contexts – when someone is genuinely trying to convince us of something. This is not the case in these decontextualized tasks that involve no interaction or in abstract problems. In fact, as soon as these logical problems can be made sense of in an argumentative context, performance improves. For instance, participants can easily understand a *modus tollens* argument when it is of use not simply to pass some test but to evaluate communicated information (see Thompson et al. 2005b); the production of valid *modus tollens* arguments in argumentative contexts is also “surprisingly common” (Pennington & Hastie 1993, p. 155).

While students of reasoning focus on logical fallacies, other scholars have turned to the study of the fallacies of argumentation. Unlike logical fallacies, fallacies of argumentation come in degrees: Depending on their content and context, they can be more or less fallacious. For instance, a *slippery-slope fallacy* (where a claim is criticized for being a step on a slope that ends up with a blatant mistake) is in fact valid to the extent that, having made the first step on the slope, it is probable that one will continue all the way down (Corner et al. 2006).

Various experiments have shown that participants are generally able to spot other argumentative fallacies (Hahn & Oaksford 2007, experiment 3; Neuman 2003; Neuman et al. 2006; Weinstock et al. 2004; see also Corner & Hahn 2009). Not only do they spot them, but they tend to react appropriately: rejecting them when they are indeed fallacious, or being convinced to the degree that they are well grounded (Corner et al. 2006; Hahn & Oaksford 2007; Hahn et al. 2005; Oaksford &

Hahn 2004; Rips 2002). When researchers have studied other skills specific to argumentation, performance has proved to be satisfactory. Thus, participants are able to recognize the macrostructure of arguments (Ricco 2003), to follow the commitments of different speakers (Rips 1998), and to attribute the burden of proof appropriately (Bailenson & Rips 1996; see also Rips 1998, experiment 3). On the whole, the results reviewed in this section demonstrate that people are good at evaluating arguments both at the level of individual inferences and at the level of whole discussions.

## 2.2. Producing arguments

The first studies that systematically investigated argument production used the following methodology.<sup>3</sup> Participants were asked to think about a given topic, such as “Would restoring the military draft significantly increase America’s ability to influence world events?” (Perkins 1985) or “What are the causes of school failure?” (Kuhn 1991). After being left to think for a few minutes, they had to state and defend their view to the experimenter. The conclusions of these studies were quite bleak and highlighted three main flaws. The first is that people resort to mere explanations (“make sense” causal theories) instead of relying on genuine evidence (data) to support their views. However, later research has shown that this is mostly an artifact of the lack of evidence available to the participants: When evidence is made available, participants will favor it (in both production and evaluation) (Brem & Rips 2000; see also Hagler & Brem 2008; Sá et al. 2005). A second flaw noted by Perkins and Kuhn is the relative superficiality of the arguments used by participants. This can be explained by a feature of the tasks: Unlike in a real debate, the experimenter didn’t challenge the arguments of the participants, however weak they were. In a normal argumentative setting, a good argument is an argument that is not refuted. As long as they are not challenged, it makes sense to be satisfied with seemingly superficial arguments. On the other hand, people should be able to generate better arguments when engaged in a real debate. This is exactly what Kuhn and her colleagues observed: Participants who had to debate on a given topic showed a significant improvement in the quality of the arguments they used afterwards (Kuhn et al. 1997; for similar results with analogical reasoning, see Blanchette & Dunbar 2001).

The third flaw, according to Perkins and Kuhn, is the most relevant one here. Participants had generally failed to anticipate counterarguments and generate rebuttals. For these two authors, and indeed the critical thinking tradition, this is a very serious failing. Seen from an argumentative perspective, however, this may not be a simple flaw but rather a feature of argumentation that contributes to its effectiveness in fulfilling its function. If one’s goal is to convince others, one should be looking first and foremost for supportive arguments. Looking for counterarguments against one’s own claims may be part of a more sophisticated and effortful argumentative strategy geared to anticipating the interlocutor’s response, but, in the experimental setting, there was no back-and-forth to encourage such an extra effort (and participants knew not to expect such a back-and-forth). If this is a correct explanation of what need not be a flaw after all, then the

difficulty that people seem to have in coming up with counterarguments should be easily overcome by having them challenge someone else’s claims rather than defending their own. Indeed, when mock jurors were asked to reach a verdict and were then presented with an alternative verdict, nearly all of them were able to find counterarguments against it (Kuhn et al. 1994). In another experiment, *all* participants were able to find counterarguments against a claim (which was not theirs) and to do so very quickly (Shaw 1996).

When people have looked at reasoning performance in felicitous argumentative settings, they have observed good results. Resnick and her colleagues (1993) created groups of three participants who disagreed on a given issue. Analyzing the debates, the researchers were “impressed by the coherence of the reasoning displayed. Participants . . . appear to build complex arguments and attack structure. People appear to be capable of recognizing these structures and of effectively attacking their individual components as well as the argument as a whole” (pp. 362–63; see also Blum-Kulka et al. 2002; Hagler & Brem 2008; Stein et al. 1997; Stein et al. 1996). It is worth noting that a strikingly similar pattern emerges from developmental studies (see Mercier, in press b).

To sum up, people can be skilled arguers, producing and evaluating arguments felicitously. This good performance stands in sharp contrast with the abysmal results found in other, nonargumentative, settings, a contrast made particularly clear by the comparison between individual and group performance.

## 2.3. Group reasoning

If people are skilled at both producing and evaluating arguments, and if these skills are displayed most easily in argumentative settings, then debates should be especially conducive to good reasoning performance. Many types of tasks have been studied in group settings, with very mixed results (for recent reviews,<sup>4</sup> see Kerr & Tindale 2004; Kerr et al. 1996). The most relevant findings here are those pertaining to logical or, more generally, intellectual tasks “for which there exists a demonstrably correct answer within a verbal or mathematical conceptual system” (Laughlin & Ellis 1986, p. 177). In experiments involving this kind of task, participants in the experimental condition typically begin by solving problems individually (pretest), then solve the same problems in groups of four or five members (test), and then solve them individually again (posttest), to ensure that any improvement does not come simply from following other group members. Their performance is compared with those of a control group of participants who take the same tests but always individually. Intellectual tasks allow for a direct comparison with results from the individual reasoning literature, and the results are unambiguous. The dominant scheme (Davis 1973) is *truth wins*, meaning that, as soon as one participant has understood the problem, she will be able to convince the whole group that her solution is correct (Bonner et al. 2002; Laughlin & Ellis 1986; Stasson et al. 1991).<sup>5</sup> This can lead to big improvements in performance. Some experiments using the Wason selection task dramatically illustrate this phenomenon (Moshman & Geil 1998; see also Augustinova 2008; Maciejovsky & Budescu 2007). The Wason selection task is the most widely used task in reasoning,

and the performance of participants is generally very poor, hovering around 10% of correct answers (Evans 1989; Evans et al. 1993; Johnson-Laird & Wason 1970). However, when participants had to solve the task in groups, they reached the level of 80% of correct answers.

Several challenges can be leveled against this interpretation of the data. It could be suggested that the person who has the correct solution simply points it out to the others, who immediately accept it without argument, perhaps because they have recognized this person as the “smartest” (Oaksford et al. 1999). The transcripts of the experiments show that this is not the case: Most participants are willing to change their mind only once they have been thoroughly convinced that their initial answer was wrong (e.g., see Moshman & Geil 1998; Trognon 1993). More generally, many experiments have shown that debates are essential to any improvement of performance in group settings (for a review and some new data, see Schulz-Hardt et al. 2006; for similar evidence in the development and education literature, see Mercier, in press b). Moreover, in these contexts, participants decide that someone is smart based on the strength and relevance of her arguments and not the other way around (Littlepage & Mueller 1997). Indeed, it would be very hard to tell who is “smart” in such groups – even if general intelligence were easily perceptible, it correlates only .33 with success in the Wason selection task (Stanovich & West 1998). Finally, in many cases, no single participant had the correct answer to begin with. Several participants may be partly wrong and partly right, but the group will collectively be able to retain only the correct parts and thus converge on the right answer. This leads to the *assembly bonus effect*, in which the performance of the group is better than that of its best member (Blinder & Morgan 2000; Laughlin et al. 2002; 2003; 2006; Lombardelli et al. 2005; Michaelsen et al. 1989; Sniezek & Henry 1989; Stasson et al. 1991; Tindale & Sheffey 2002). Once again there is a striking convergence here, with the developmental literature showing how groups – even when no member had the correct answer initially – can facilitate learning and comprehension of a wide variety of problems (Mercier, in press b).

According to another counterargument, people are simply more motivated, generally, when they are in groups (Oaksford et al. 1999). This is not so.<sup>6</sup> On the contrary, “the ubiquitous finding across many decades of research (e.g., see Hill 1982; Steiner 1972) is that groups usually fall short of reasonable potential productivity baselines” (Kerr & Tindale 2004, p. 625). Moreover, other types of motivation have no such beneficial effect on reasoning. By and large, monetary incentives, even substantial ones, fail to improve performance in reasoning and decision-making tasks (Ariely et al., 2009; Bonner & Sprinkle 2002; Bonner et al. 2000; Camerer & Hogarth 1999; and, in the specific case of the Wason selection task, see Johnson-Laird & Byrne 2002; Jones & Sugden, 2001). Thus, not any incentive will do: Group settings have a motivational power to which reasoning responds specifically.<sup>7</sup>

The argumentative theory also helps predict what will happen in nonoptimal group settings. If all group members share an opinion, a debate should not arise spontaneously. However, in many experimental and institutional settings (juries, committees), people are forced to discuss, even if they already agree. When all group members agree on a certain view, each of them can find

arguments in its favor. These arguments will not be critically examined, let alone refuted, thus providing other group members with additional reasons to hold that view. The result should be a strengthening of the opinions held by the group (for a review, see Sunstein 2002; for a recent illustration, see Hinsz et al. 2008). Contra Sunstein’s *law* of group polarization, it is important to bear in mind that this result is specific to artificial contexts in which people debate even though they tend to agree in the first place. When group members disagree, discussions often lead to *depolarization* (Kogan & Wallach 1966; Vinokur & Burnstein 1978). In both cases, the behavior of the group can be predicted on the basis of the direction and strength of the arguments accessible to group members, as demonstrated by research carried out in the framework of the Persuasive Argument Theory (Vinokur 1971), which ties up with the prediction of the present framework (Ebbesen & Bowers 1974; Isenberg 1986; Kaplan & Miller 1977; Madsen 1978).

The research reviewed in this section shows that people are skilled arguers: They can use reasoning both to evaluate and to produce arguments. This good performance offers a striking contrast with the poor results obtained in abstract reasoning tasks. Finally, the improvement in performance observed in argumentative settings confirms that reasoning is at its best in these contexts. We will now explore in more depth a phenomenon already mentioned in this section: the confirmation bias.

### 3. The confirmation bias: A flaw of reasoning or a feature of argument production?

The confirmation bias consists in the “seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand” (Nickerson 1998, p. 175). It is one of the most studied biases in psychology (for review, see Nickerson 1998). While there is some individual variation, it seems that everybody is affected to some degree, irrespective of factors like general intelligence or open mindedness (Stanovich & West 2007; 2008a; 2008b). For standard theories of reasoning, the confirmation bias is no more than a *flaw of reasoning*. For the argumentative theory, however, it is a consequence of the function of reasoning and hence a *feature* of reasoning when used for the production of arguments.

In fact, we suggest, the label *confirmation bias* has been applied to two distinct types of case, both characterized by a failure to look for counterevidence or counterarguments to an existing belief, both consistent with the argumentative approach but brought about in different ways. In cases that deserve the label of confirmation bias, people are trying to convince others. They are typically looking for arguments and evidence to confirm their own claim, and ignoring negative arguments and evidence unless they anticipate having to rebut them. While this may be seen as a bias from a normative epistemological perspective, it clearly serves the goal of convincing others. In another type of case, we are dealing not with biased reasoning but with an absence of reasoning proper. Such an absence of reasoning is to be expected when people already hold some belief on the basis of perception, memory, or intuitive inference, and do not have to argue for it. Say, I believe that my keys are in my trousers because that is where I remember putting them. Time has



passed, and they could now be in my jacket, for example. However, unless I have some positive reason to think otherwise, I just assume that they are still in my trousers, and I don't even make the inference (which, if I am right, would be valid) that they are not in my jacket or any of the other places where, in principle, they might be. In such cases, people typically draw positive rather than negative inferences from their previous beliefs. These positive inferences are generally more relevant to testing these beliefs. For instance, I am more likely to get conclusive evidence that I was right or wrong by looking for my keys in my trousers rather than in my jacket (even if they turn out not to be in my jacket, I might still be wrong in thinking that they are in my trousers). We spontaneously derive positive consequences from our intuitive beliefs. This is just a trusting use of our beliefs, not a confirmation bias (see Klayman & Ha 1987).

The theory we are proposing makes three broad predictions. The first is that the genuine confirmation bias (as opposed to straightforward trust in one's intuitive beliefs and their positive consequences) should occur only in argumentative situations. The second is that it should occur only in the *production* of arguments. The rationale for a confirmation bias in the production of arguments to support a given claim does not extend to the *evaluation* of arguments by an audience that is just aiming to be well informed. The third prediction is that the confirmation bias in the production of arguments is not a bias in favor of confirmation in general and against disconfirmation in general: It is a bias in favor of confirming one's own claims, which should be naturally complemented by a bias in favor of disconfirming opposing claims and counterarguments.

### 3.1. Hypothesis testing: No reasoning, no reasoning bias

One of the areas in which the confirmation bias has been most thoroughly studied is that of hypothesis testing, often using Wason's rule discovery task (Wason 1960). In this task, participants are told that the experimenter has in mind a rule for generating number triples and that they have to discover it. The experimenter starts by giving participants a triple that conforms to the rule (2, 4, 6). Participants can then think of a hypothesis about the rule and test it by proposing a triple of their own choice. The experimenter says whether or not this triple conforms to the rule. Participants can repeat the procedure until they feel ready to put forward their hypothesis about the rule. The experimenter tells them whether or not their hypothesis is true. If it is not, they can try again or give up.

Participants overwhelmingly propose triples that fit with the hypothesis they have in mind. For instance, if a participant has formed the hypothesis "three even numbers in ascending order," she might try 8, 10, 12. As argued by Klayman and Ha (1987), such an answer corresponds to a "positive test strategy" of a type that would be quite effective in most cases. This strategy is not adopted in a reflective manner, but is rather, we suggest, the intuitive way to exploit one's intuitive hypotheses, as when we check that our keys are where we believe we left them as opposed to checking that they are not where it follows from our belief that they should not be. What we see here, then, is a sound heuristic rather than a bias.

This heuristic misleads participants in this case only because of some very peculiar (and expressly designed) features of the task. What is really striking is the failure of attempts to get participants to reason in order to correct their ineffective approach. It has been shown that, even when instructed to try to falsify the hypotheses they generate, fewer than one participant in ten is able to do so (Poletiek 1996; Tweney et al. 1980). Since the hypotheses are generated by the participants themselves, this is what we should expect in the current framework: The situation is not an argumentative one and does not activate reasoning. However, if a hypothesis is presented as coming from someone else, it seems that more participants will try to falsify it and will give it up much more readily in favor of another hypothesis (Cowley & Byrne 2005). The same applies if the hypothesis is generated by a minority member in a group setting (Butera et al. 1992). Thus, falsification is accessible provided that the situation encourages participants to argue against a hypothesis that is not their own.

### 3.2. The Wason selection task

A similar interpretation can be used to account for results obtained with the Wason selection task (Wason 1966). In this task, participants are given a rule describing four cards. In the original version, the cards have a number on one side and a letter on the other, although only one side is visible – they might see, for instance, 4, E, 7, and K. The rule might read, "If there is a vowel on one side, then there is an even number on the other side." The task is to say what cards need to be turned over to determine whether the rule is true. In this task, too, it is useful to distinguish the effects of intuitive mechanisms from those of reasoning proper (as has long been suggested by Wason & Evans 1975). Intuitive mechanisms involved in understanding utterances will draw the participants' attention to the cards that are made most relevant by the rule and the context (Giroto et al. 2001; Sperber et al. 1995). In the standard case, these will simply be the cards mentioned in the rule (the vowel E and the even number 4), as opposed to those that would yield the correct answer (the E and the 7). Given that the 4 can only confirm the rule but not falsify it, the behavior of participants who select this card could be interpreted as showing a confirmation bias. However, as first discovered by Evans (Evans & Lynch 1973), the simple addition of a negation in the rule ("if there is a vowel on one side, then there is *not* an even number on the other side") leaves the answers unchanged (the E and 4 are still made relevant), but in this case these cards correspond to the correct, falsifying, response. So these intuitive mechanisms are not intrinsically linked to either confirmation or falsification: They just happen to point to cards that in some cases might confirm the rule and, in other cases, might falsify it.

Confirmation bias does occur in the selection task but at another level. Once the participants' attention has been drawn to some of the cards, and they have arrived at an intuitive answer to the question, reasoning is used not to evaluate and correct their initial intuition but to find justifications for it (Evans 1996; Lucas & Ball 2005; Roberts & Newton 2001). This is a genuine confirmation bias. As with hypothesis testing, this does not mean that participants are

simply unable to understand the task or to try to falsify the rule – only that an appropriate argumentative motivation is lacking. That participants can understand the task is shown by the good performance in group settings, as mentioned earlier. Participants should also be able to try to falsify the rule when their first intuition is that the rule is false and they want to *prove it wrong*. Researchers have used rules such as “all members of group A are Y,” where Y is a negative or positive stereotype (Dawson et al. 2002). Participants who were most motivated to prove the rule wrong – those belonging to group A when Y was negative – were able to produce more than 50% of correct answers, whereas participants from all the other conditions (groups other than A and/or positive stereotype) remained under 20%.

### 3.3. Categorical syllogisms

Categorical syllogisms are one of the most studied types of reasoning. Here is a typical example: “No C are B; All B are A; therefore some A are not C.” Although they are solvable by very simple programs (e.g., see Geurts 2003), syllogisms can be very hard to figure out – the one just offered by way of illustration, for instance, is solved by less than 10% of participants (Chater & Oaksford 1999). In terms of the mental model theory, what the participants are doing is constructing a model of the premises and deriving a possible conclusion from it (Evans et al. 1999). This constitutes the participants’ initial intuition. To correctly solve the problem, participants should then try to construct counterexamples to this initial conclusion. But this would mean trying to falsify their own conclusion. The present theory predicts that they will not do so spontaneously. And indeed, “any search for counterexample models is weak . . . participants are basing their conclusions on the first model that occurs to them” (Evans et al. 1999, p. 1505; see also Klauer et al. 2000; Newstead et al. 1999).

Again, we suggest, this should not be interpreted as revealing a lack of ability but only a lack of motivation. When participants want to prove a conclusion wrong, they will find ways to falsify it.

This happens with normal conclusions presented by someone else (Sacco & Bucciarelli 2008) or when participants are faced with so-called unbelievable conclusions such as “All fish are trout.” In this case, they will try to prove that the premises lead to the logical opposite of the conclusion (“Not all fish are trout”) (Klauer et al. 2000). Given that falsification leads to better answers on these tasks, this explains why participants actually perform much better when the conclusion is unbelievable (e.g., see Evans et al. 1983). It is not that they reason *more* in this case – they spend as much time trying to solve problems with believable conclusions as with unbelievable ones (Thompson et al. 2003). It is just that the direction reasoning takes is mostly determined by the participants’ initial intuitions. If they have arrived at the conclusion themselves, or if they agree with it, they try to confirm it. If they disagree with it, they try to prove it wrong. In all cases, what they do is try to confirm their initial intuition.

### 3.4. Rehabilitating the confirmation bias

In all three cases just reviewed – hypothesis testing, the Wason selection task, and syllogistic reasoning – a

similar pattern can be observed. Participants have intuitions that lead them towards certain answers. If reasoning is used at all, it is mostly used to confirm these initial intuitions. This is exactly what one should expect of an argumentative skill, and so these results bolster our claim that the main function of reasoning is argumentative. By contrast, if people were easily able to abstract from this bias, or if they were subject to it only in argumentative settings, then this would constitute evidence against the present theory.

According to a more standard explanation of the confirmation bias, it is an effect of limitations in cognitive resources and in particular in working memory (e.g., Johnson-Laird 2006). But it is hard to reconcile this explanation with the fact that people are very good at falsifying propositions when they are inclined to disagree with them. In those cases, people are not held back by limited resources even though the tasks are not cognitively easier.

However, the idea that the confirmation bias is a normal feature of reasoning that plays a role in the production of arguments may seem surprising in light of the poor outcomes it has been claimed to cause. Conservatism in science is one example (see Nickerson 1998 and references therein). Another is the related phenomenon of groupthink, which has been held responsible for many disasters, from the Bay of Pigs fiasco (Janis 1982) to the tragedy of the Challenger shuttle (Esser & Lindorfer 1989; Moorhead et al. 1991) (for review, see Esser 1998). In such cases, reasoning tends not to be used in its normal context: that is, the resolution of a disagreement through discussion. When one is alone or with people who hold similar views, one’s arguments will not be critically evaluated. This is when the confirmation bias is most likely to lead to poor outcomes. However, when reasoning is used in a more felicitous context – that is, in arguments among people who disagree but have a common interest in the truth – the confirmation bias contributes to an efficient form of *division of cognitive labor*.

When a group has to solve a problem, it is much more efficient if each individual looks mostly for arguments supporting a given solution. They can then present these arguments to the group, to be tested by the other members. This method will work as long as people can be swayed by good arguments, and the results reviewed in section 2 show that this is generally the case. This joint dialogic approach is much more efficient than one where each individual on his or her own has to examine all possible solutions carefully.<sup>8</sup> The advantages of the confirmation bias are even more obvious given that each participant in a discussion is often in a better position to look for arguments in favor of his or her favored solution (situations of asymmetrical information). So group discussions provide a much more efficient way of holding the confirmation bias in check. By contrast, the teaching of critical thinking skills, which is supposed to help us overcome the bias on a purely individual basis, does not seem to yield very good results (Ritchart & Perkins 2005; Willingham 2008).

For the confirmation bias to play an optimal role in discussions and group performance, it should be active only in the production of arguments and not in their evaluation. Of course, in the back-and-forth of a discussion, the production of one’s own arguments and the evaluation of those of the interlocutor may interfere with each other, making it hard to properly assess the two processes

independently. Still, the evidence reviewed in section 2.1 on the understanding of arguments strongly suggests that people tend to be more objective in evaluation than in production. If this were not the case, the success of group reasoning reviewed in section 2.3 would be very hard to explain.

#### 4. Proactive reasoning in belief formation

According to the argumentative theory, reasoning is most naturally used in the context of an exchange of arguments during a discussion. But people can also be proactive and anticipate situations in which they might have to argue to convince others that their claims are true or that their actions are justified. We would say that much reasoning anticipates the need to argue. In this section, we will show that work on motivated reasoning can be usefully reinterpreted in this perspective, and, in the next section, we will show that the same applies to work on reason-based choice.

Many of our beliefs are likely to remain unchallenged because they are relevant only to ourselves and we don't share them or because they are uncontroversial among the people we interact with or because we have sufficient authority to be trusted when we assert them. While we think of most of our beliefs – to the extent that we think about them at all – not as beliefs but just as pieces of knowledge, we are also aware that some of them are unlikely to be universally shared, or to be accepted on trust just because we express them. When we pay attention to the contentious nature of these beliefs, we typically think of them as opinions. Opinions are likely to be challenged and may have to be defended. It makes sense to look for arguments for our opinions before we find ourselves called upon to state them. If the search for arguments is successful, we will be ready. If not, then perhaps it might be better to adopt a weaker position, one that is easier to defend. Such uses of reasoning have been intensively studied under the name of *motivated reasoning*<sup>9</sup> (Kunda 1990; see also Kruglanski & Freund 1983; Pyszczynski & Greenberg 1987; for a recent review, see Molden & Higgins 2005).

##### 4.1. Motivated reasoning

A series of experiments by Ditto and his colleagues, involving reasoning in the context of a fake medical result, illustrate the notion of motivated reasoning (Ditto & Lopez 1992; Ditto et al. 1998; 2003). Participants had to put some saliva on a strip of paper and were told that, if the strip changed color or did not change color, depending on the condition, this would be an indication of an unhealthy enzyme deficiency. Participants, being motivated to believe they were healthy, tried to garner arguments for this belief. In one version of the experiment, participants were told the rate of false positives, which varied across conditions. The use they made of this information reflects motivated reasoning. When the rate of false positives was high, participants who were motivated to reject the conclusion used it to undermine the validity of the test. This same high rate of false positives was discounted by participants who were motivated to accept the conclusion. In another version of the experiment,

participants were asked to mention events in their medical history that could have affected the results of the test, which gave them an opportunity to discount these results. Participants motivated to reject the conclusion listed more such events, and the number of events listed was negatively correlated with the evaluation of the test. In these experiments, the very fact that the participant's health is being tested indicates that it cannot be taken for granted. The reliability of the test itself is being discussed. This experiment, and many others to be reviewed in this article, demonstrate also that motivated reasoning is not mere wishful thinking (a form of thinking that, if it were common, would in any case be quite deleterious to fitness and would not be coherent with the present theory). If desires did directly affect beliefs in this way, then participants would simply ignore or dismiss the test. Instead, what they do is look for evidence and arguments to show that they are healthy or at least for reasons to question the value of the test.

Other studies have demonstrated the use of motivated reasoning to support various beliefs that others might challenge. Participants dig in and occasionally alter their memories to preserve a positive view of themselves (Dunning et al. 1989; Ross et al. 1981; Sanitioso et al. 1990). They modify their causal theories to defend some favored belief (Kunda 1987). When they are told the outcome of a game on which they had made a bet, they use events in the game to explain why they should have won when they lost (Gilovich 1983). Political experts use similar strategies to explain away their failed predictions and bolster their theories (Tetlock 1998). Reviewers fall prey to motivated reasoning and look for flaws in a paper in order to justify its rejection when they don't agree with its conclusions (Koehler 1993; Mahoney 1977). In economic settings, people use information flexibly so as to be able to justify their preferred conclusions or arrive at the decision they favor (Boiney et al. 1997; Hsee 1995; 1996a; Schweitzer & Hsee 2002).

All these experiments demonstrate that people sometimes look for reasons to justify an opinion they are eager to uphold. From an argumentative perspective, they do this not to convince themselves of the truth of their opinion but to be ready to meet the challenges of others. If they find themselves unprepared to meet such challenges, they may become reluctant to express an opinion they are unable to defend and less favorable to the opinion itself, but this is an indirect individual effect of an effort that is aimed at others. In a classical framework, where reasoning is seen as geared to achieving epistemic benefits, the fact that it may be used to justify an opinion already held is hard to explain, especially since, as we will now show, motivated reasoning can have dire epistemic consequences.

##### 4.2. Consequences of motivated reasoning

**4.2.1. Biased evaluation and attitude polarization.** In a landmark experiment, Lord and colleagues (1979) asked participants who had been previously selected as being either defenders or opponents of the death penalty to evaluate studies relating to its efficiency as a deterrent. The studies given to the participants had different conclusions: While one seemed to show that the death penalty had a significant deterrent effect, the other

yielded the opposite result. Even though the methodologies of the two studies were almost identical, the studies that yielded a conclusion not in line with the participants' opinions were consistently rated as having been much more poorly conducted. In this case, participants used reasoning not so much to assess the studies objectively as to confirm their initial views by finding either flaws or strengths in similar studies, depending on their conclusion. This phenomenon is known as *biased assimilation* or *biased evaluation*. This second description is somewhat misleading. In this experiment – and the many related experiments that have followed it – participants are indeed asked to evaluate an argument. However, what they do is mostly *produce* arguments to support or rebut the argument they are evaluating, depending on whether they agree with its conclusion or not. Participants are not trying to form an opinion: They already have one. Their goal is argumentative rather than epistemic, and it ends up being pursued at the expense of epistemic soundness. That participants engage in this biased search for arguments even when their task is to evaluate an argument has been demonstrated by the experiments we now describe.

Several other experiments have studied the way people evaluate arguments depending on whether they agree or disagree with the conclusions. When people disagree with the conclusion of an argument, they often spend more time evaluating it (Edwards & Smith 1996). This asymmetry arises from the trivial fact that rejecting what we are told generally requires some justification, whereas accepting it does not. Moreover, the time spent on these arguments is mostly devoted to finding counterarguments (Edwards & Smith 1996; see also Brock 1967; Cacioppo & Petty 1979; Eagly et al. 2000). Participants tend to comb through arguments for flaws and end up finding some, whether they are problems with the design of a scientific study (Klaczynski & Gordon 1996b; Klaczynski & Narasimham 1998; Klaczynski & Robinson 2000), issues with a piece of statistical reasoning (Klaczynski & Gordon 1996a; Klaczynski & Lavalley 2005; Klaczynski et al. 1997), or argumentative fallacies (Klaczynski 1997). In all these cases, motivated reasoning leads to a biased assessment: Arguments with unfavored conclusions are rated as less sound and less persuasive than arguments with favored conclusions.

Sometimes the evaluation of an argument is biased to the point where it has an opposite effect to the one intended by the arguer: On reading an argument with a counterattitudinal conclusion (one that goes against their own beliefs or preferences), interlocutors may find so many flaws and counterarguments that their initial unfavorable attitude is in fact strengthened. This is the phenomenon of attitude polarization, which has been studied extensively since it was first demonstrated by (Lord et al. 1979; see also Greenwald 1969; Pomerantz et al. 1995).<sup>10</sup> Taber and Lodge (2006) have demonstrated that, in the domain of politics, attitude polarization is most easily observed in participants who are most knowledgeable (see also Braman 2009; Redlawsk 2002). Their knowledge makes it possible for these participants to find more counterarguments, leading to more biased evaluations.

#### 4.2.2. Polarization, bolstering, and overconfidence.

Attitude polarization can also occur in simpler circumstances. Merely thinking about an object may be enough

to strengthen attitudes towards it (*polarization*). This phenomenon has been repeatedly demonstrated. Sadler and Tesser (1973) had participants listen to a recording of a very pleasant-sounding or unpleasant-sounding individual. They then had to give their opinion of this individual, either after having to think about him or her or after performing a distraction task. As expected, the opinions were more extreme (in both directions) when participants had to think about the individual. Tesser and Conlee (1975) showed that polarization increases with the time spent thinking about an item, and Jellison and Mills (1969) showed that it increases with the motivation to think. As in the case of polarization following biased evaluation, such polarization occurs only when participants are knowledgeable (Tesser & Leone 1977; see also Millar & Tesser 1986). And the effect can be mitigated by providing a reality check: The simple presence of the target object will dramatically decrease polarization (Tesser 1976).

Some later experiments used a slightly different methodology (Chaiken & Yates 1985; Liberman & Chaiken 1991). Instead of simply thinking about the target object, participants had to write a small essay about it. Not only was polarization observed in this case, but it was correlated with the direction and number of the arguments put forward in the essay. These results demonstrate that reasoning contributes to attitude polarization and strongly suggest that it may be its main factor. When people are asked to think about a given item toward which they intuitively have a positive or negative attitude, what happens, we suggest, is that they reflect less on the item itself than on how to defend their initial attitude. Many other experiments have shown that, once people have formed an attitude to a target, they will look for information that supports this attitude (a phenomenon known as *selective exposure*; see Hart et al. 2009; Smith et al. 2008) and try to put any information they are given to the same use (Bond et al. 2007; Brownstein 2003), which leads them to choose inferior alternatives (Russo et al. 2006).

According to the argumentative theory, reasoning should be even more biased once the reasoner has already stated her opinion, thereby increasing the pressure on her to justify it rather than moving away from it. This phenomenon is called *bolstering* (McGuire 1964). Thus, when participants are committed to an opinion, thinking about it will lead to a much stronger polarization (Lambert et al. 1996; Millar & Tesser 1986). *Accountability* (the need to justify one's decisions) will also increase bolstering (Tetlock et al. 1989; for review, see Lerner & Tetlock 1999).

Finally, motivated reasoning should also affect confidence. When participants think of an answer to a given question, they will be spontaneously tempted to generate reasons supporting that answer. This may then cause them to be overconfident in the answer. Koriat and his colleagues (1980) have tested this hypothesis by using general knowledge questions such as “the Sabines were part of (a) ancient India or (b) ancient Rome.” After answering the question, participants had to produce reasons relevant to their answers. Some participants were asked to generate reasons supporting their answer, while others were asked for reasons against it. The results for people who were explicitly asked to generate reasons supporting their answer were no different from those in a control condition where no reasons were asked for. This suggests that

thinking of reasons to support their answer is what people do spontaneously anyhow when they regard their answer not as an obvious piece of knowledge but as an opinion that might be challenged. By contrast, participants in the other group were much less overconfident. Having to think of arguments against their answer enabled them to see its limitations – something they would not do on their own (for replications and extensions to the phenomenon of hindsight bias and the fundamental attribution error, see Arkes et al. 1988; Davies 1992; Griffin & Dunning 1990; Hirt & Markman 1995; Hoch 1985; Yates et al. 1992). It is then easy to see that overconfidence would also be reduced by having participants discuss their answers with people who favor different conclusions.

**4.1.3. Belief perseverance.** Motivated reasoning can also be used to hang on to beliefs even when they have been proved to be ill-founded. This phenomenon, known as *belief perseverance*, is “one of social psychology’s most reliable phenomena” (Guenther & Alicke 2008, p. 706; for an early demonstration, see Ross et al. 1975). The involvement of motivated reasoning in this effect can be demonstrated by providing participants with evidence both for and against a favored belief. If belief perseverance were a simple result of some degree of psychological inertia, then the first evidence presented should be the most influential, whether it supports or disconfirms the favored belief. On the other hand, if evidence can be used selectively, then only evidence supporting the favored belief should be retained, regardless of the order of presentation. Guenther and Alicke 2008 tested this hypothesis in the following way. Participants first had to perform a simple perceptual task. This task, however, was described as testing for “mental acuity,” a made-up construct that was supposed to be related to general intelligence, making the results of the test highly relevant to participant’s self-esteem. Participants were then given positive or negative feedback, but a few minutes later they were told that the feedback was actually bogus and the real aim of the experiment was explained. At three different points, the participants also had to evaluate their performance: right after the task, after the feedback, and after the debriefing. In line with previous results, the participants who had received positive feedback showed a classic belief-perseverance effect and discounted the debriefing, which allowed them to preserve a positive view of their performance. By contrast, those who had received negative feedback did the opposite: They took the debriefing fully into account, which allowed them to reject the negative feedback and restore a positive view of themselves. This strongly suggests that belief perseverance of the type just described is an instance of motivated reasoning (for applications to the domain of political beliefs, see Prasad et al. 2009).<sup>11</sup>

**4.1.4. Violation of moral norms.** The results reviewed so far have shown that motivated reasoning can lead to poor epistemic outcomes. We will now see that our ability to “find or make a reason for everything one has a mind to do” (Franklin 1799) can also allow us to violate our moral intuitions and behave unfairly. In a recent experiment, Valdesolo and DeSteno (2008) have demonstrated the role reasoning can play in maintaining moral hypocrisy (when we judge someone else’s action by

using tougher moral criteria than we use to judge our own actions). Here is the basic setup. On arriving at the laboratory, participants were told that they would be performing one of two tasks: a short and fun task or a long and hard task. Moreover, they were given the possibility of choosing which task they would be performing, knowing that the other task would be assigned to another participant. They also had the option of letting a computer choose at random how the tasks would be distributed. Once they were done assigning the tasks, participants had to rate how fair they had been. Other participants, instead of having to make the assignment themselves, were at the receiving end of the allocation and had no choice whatsoever; they had to rate the fairness of the participant who had done the allocation, knowing the exact conditions under which this had been done. It is then possible to compare the fairness ratings of participants who have assigned themselves the easy task with the ratings of those who have been assigned the hard task. The difference between these two ratings is a mark of moral hypocrisy. The authors then hypothesized that reasoning, since it allows participants to find excuses for their behavior, was responsible for this hypocrisy. They tested this hypothesis by replicating the above conditions with a twist: The fairness judgments were made under cognitive load, which made reasoning close to impossible. This had the predicted result: Without the opportunity to reason, the ratings were identical and showed no hint of hypocrisy.

This experiment is just one illustration of a more general phenomenon. Reasoning is often used to find justifications for performing actions that are otherwise felt to be unfair or immoral (Bandura 1990; Bandura et al. 1996; Bersoff 1999; Crandall & Eshleman 2003; Dana et al. 2007; Diekmann et al. 1997; Haidt 2001; Mazar et al. 2008; Moore et al. 2008; Snyder et al. 1979; for children, see Gummerum et al. 2008). Such uses of reasoning can have dire consequences. Perpetrators of crimes will be tempted to “blame the victim” or find other excuses to mitigate the effects of violating their moral intuitions (Ryan 1971; for a review, see Hafer & Begue 2005), which can in turn make it easier to commit new crimes (Baumeister 1997). This view of reasoning dovetails with recent theories of moral reasoning that see it mostly as a tool for communication and persuasion (Gibbard 1990; Haidt 2001; Haidt & Bjorklund 2007).

These results raise a problem for the classical view of reasoning. In all these cases, reasoning does not lead to more accurate beliefs about an object, or to better estimates of the correctness of one’s answer, or to superior moral judgments. Instead, by looking only for supporting arguments, reasoning strengthens people’s opinions, distorts their estimates, and allows them to get away with violations of their own moral intuitions. In these cases, epistemic or moral goals are not well served by reasoning. By contrast, argumentative goals are: People are better able to support their positions or to justify their moral judgments.

## 5. Proactive reasoning in decision making

In the previous section, we argued that much reasoning is done in anticipation of situations where an opinion might have to be defended, and we suggested that work on motivated reasoning can be fruitfully reinterpreted in

this light. It is not just opinions that may have to be defended: People may also have to put forward arguments to defend their decisions and actions, and they may reason proactively to that end. We want to argue that this is the main role of reasoning in decision making. This claim stands in sharp contrast to the classical view that reasoning about possible options and weighing up their pros and cons is the most reliable way – if not the only reliable way – to arrive at sound decisions (Janis & Mann 1977; Kahneman 2003; Simon 1955). This classical view has in any case been vigorously challenged in much recent research. Some argue that the best decisions are based on intuition and made in split seconds (e.g., see Klein 1998), a view rendered popular by Gladwell (2005). Others maintain that the solution lies with the unconscious and advise us to “sleep on it” (Claxton 1997; Dijksterhuis 2004; Dijksterhuis & van Olden 2006; Dijksterhuis et al. 2006b). We briefly review these challenges to the classical view before considering the substantial literature on reason-based choice and interpreting it in the light of the argumentative theory of reasoning.

### 5.1. To what extent does reasoning help in deciding?

In an initial series of studies, Wilson and his colleagues looked at the effect of reasoning on the consistency between attitudes and behavior (for review, see Wilson et al. 1989a; see also Koole et al. 2001; Millar & Tesser 1989; Sengupta & Fitzsimons 2000; 2004; Wilson & LaFleur 1995; Wilson et al. 1984; 1989b). The basic paradigm is as follows: Participants are asked to state their attitude to a given object. In one condition, they have to provide reasons for these attitudes. It has been consistently observed that attitudes based on reasons were much less predictive of future behaviors (and often not predictive at all) than were attitudes stated without recourse to reasons. This lack of correlation between attitude and behavior resulting from too much reasoning can even lead participants to form intransitive preferences (Lee et al. 2008).

Using similar paradigms in which some participants are asked for reasons, it was found that providing reasons led participants to choose items that they were later less satisfied with (Wilson et al. 1993) or that were less in line with the ratings of experts (McMackin & Slovic 2000; Wilson & Schooler 1991). Participants got worse at predicting the results of basketball games (Halberstadt & Levine 1999). People who think too much are also less likely to understand other people’s behavior (Albrechtsen et al. 2009; Ambady & Gray 2002; Ambady et al. 2000). This stream of experiments was later followed up by Dijksterhuis and his colleagues, who introduced a modified paradigm. Here, participants are given lists of features describing different items (such as flats and cars) designed in such a way that some items have more positive features. In the baseline condition, participants had to say which item they preferred immediately after they had been exposed to these features. In the conscious thought condition, they were left to think about the items for a few minutes. Finally, in the unconscious thought condition, participants spent the same amount of time doing a distraction task. Across several experiments, it was found that the best performance was obtained in this last condition: Unconscious thought was superior to conscious thought (and to immediate decision) (Dijksterhuis

2004; Dijksterhuis & van Olden 2006; Dijksterhuis et al. 2006b; 2009).

However, some of Dijksterhuis’s results have proven hard to replicate (Acker 2008; Newell et al. 2009; Thorsteinson & Withrow 2009), and alternative interpretations have been proposed in some cases (Lassiter et al. 2009). In a meta-analysis of this literature, Acker (2008) observed that in only a few experiments was unconscious thought significantly superior to conscious thought, amounting to a null result when all the experiments were taken into account. Even so, there was no significant advantage of conscious thought over immediate choice. This is typically the kind of situation where, according to classical theories, reasoning should help: A new choice has to be made, with the options well delimited and the pros and cons exposed. It is therefore quite striking that reasoning (at least for a few minutes) does not bring any advantage and is sometimes inferior to intuitive, unconscious processes. Finally, studies of decision making in natural environments converge on similar conclusions: Not only are most decisions made intuitively, but when conscious decision-making strategies are used, they often result in poor outcomes (Klein 1998). In the next subsection, we explore a framework designed to explain such findings by showing that reasoning pushes people not towards the best decisions but towards decisions that are easier to justify.

### 5.2. Reason-based choice

Starting in the late 1980s, a group of leading researchers in decision making developed the framework of *reason-based choice* (for an early review, see Shafir et al. 1993). According to this theory, people often make decisions because they can find reasons to support them. These reasons will not favor the best decisions or decisions that satisfy some criterion of rationality, but decisions that can be easily justified and are less at risk of being criticized. According to the argumentative theory, this is what should happen when people are faced with decisions where they only have weak intuitions. In this case, reasoning can be used to tip the scales in favor of the choice for which reasons are most easily available. One will then at least be able to defend the decision if its outcome proves unsatisfactory.

Reason-based choice is well illustrated in a landmark article by Simonson (1989) in which he studied, in particular, the attraction effect (Huber et al. 1982; for a cross-cultural variation, see Briley et al. 2000). The *attraction effect* occurs when, given a set of two equally valuable alternatives, a third alternative is added that is just as good as another one of the first alternatives on one trait but inferior on the second trait. This addition tends to increase the rate of choice of the dominating option in a manner not warranted by rational models. Here is one example used in Simonson’s experiments. Participants had to choose between packs of beer that varied along the two dimensions of price and quality. Beer A was of lower quality than beer B but was also cheaper, and the two attributes balanced in such a way that both beers were regularly chosen in a direct comparison. However, some participants had to choose between these two beers plus beer C, which was more expensive than beer B but not better. When this beer was introduced, participants tended to pick beer B more often. It is easy to account for this finding within the framework of reason-based

choice: The poorer alternative makes the choice of the dominating one easy to justify. (“Beer B is of the same quality as but cheaper than this other beer!”) To confirm this intuition, Simonson made and tested the three following predictions: (1) a choice based on reasons should be reinforced when participants have to justify themselves, (2) a choice based on reasons will be perceived as easier to justify and less likely to be criticized, and (3) a choice based on reasons should give rise to more elaborate explanations. The results of three experiments supported these predictions. Moreover, these results also showed that participants who made choices based on reasons tended to make choices that fitted less well with their own preferences as stated before the choice was made. Finally, another set of experiments demonstrated that, when participants were able to use their intuitions more, because they were familiar with the alternatives or because the descriptions of these alternatives were more detailed, they were less prone to the attraction effect (Ratneshwar et al. 1987). Several well-known challenges to the view of humans as making rational decisions thanks to their reasoning abilities have been, or can be, reinterpreted as cases of reason-based choice.

### 5.3. What reason-based choice can explain

**5.3.1. Disjunction effect.** The *sure-thing principle* (Savage 1954) states that, when someone favors A over B if event E happens and keeps the same preference ordering if E does not happen, then her choices should not be influenced by any uncertainty about the occurrence of E. Shafir and Tversky (1992; Tversky & Shafir 1992) have recorded several violations of this principle. For instance, we can compare the reaction of participants to the following problems (Tversky & Shafir 1992):

#### *Win/lose versions*

Imagine that you have just played a game of chance that gave you a 50% chance to win \$200 and a 50% chance to lose \$100. The coin was tossed and you have either won \$200 or lost \$100. You are now offered a second identical gamble: 50% chance to win \$200 and 50% chance to lose \$100. Would you?: (a) accept the second gamble. (b) reject the second gamble. (Tversky & Shafir 1992, p. 306)

Whether they have won or lost in the first gamble, a majority of participants accept the second gamble. However, they are likely to do so for different reasons: In the win scenario, they reason that they can easily risk losing half of the \$200 they have just won. In the lose scenario, however, they might take the second gamble as an opportunity to make up for their previous loss. In these two cases, while the choice is the same, the reasons for making it are incompatible. Thus, when participants do not know what is going to be the outcome of the first bet, they have more trouble justifying the decision to accept the second gamble: The reasons seem to contradict each other. As a result, a majority of participants who do not know the result of the first gamble reject the second gamble even though they would have accepted it whatever the result of the first gamble. The authors tested this explanation further by devising a comparison that had the same properties as the one just described, except that the reasons for making the “accept” decision were the same irrespective of the outcome of the first gamble. In this case, participants made exactly the same choices

whether or not they knew the result of the first gamble (for a similar experiment with a variant of the prisoner’s dilemma, see Croson 1999).

**5.3.2. Sunk-cost fallacy.** The *sunk-cost fallacy* is the “greater tendency to continue an endeavor once an investment in money, effort, or time has been made” (Arkes & Blumer 1985, p. 124). A well-known real-life example is that of the Concorde: The British and French governments decided to keep paying for a plane that they knew would never turn a profit. Arkes and Ayton (1999) have argued that such mistakes result from an unsatisfactory use of explicit reasons such as “do not waste.” We will briefly review the evidence they presented, and add more.

First of all, Arkes and Ayton (1999) contrast the robust sunk-cost effects observed in humans (Arkes & Blumer 1985; Garland 1990; Staw 1981) with the absence of such mistakes among animals.<sup>12</sup> They also point out that children do not seem prone to this error (for more recent, convergent evidence, see Klaczynski & Cottrell 2004; Morsanyi & Handley 2008). If reasoning were not the cause of this phenomenon but the cure for it, the opposite would be expected. Finally, some experiments have varied the availability of justifications – a factor that should not be relevant for standard models of decision making. Thus, when participants can justify the waste, they are less likely to be trapped by sunk costs (Soman & Cheema 2001). By contrast, when participants find it harder to justify changing their course of actions, they are more likely to commit the fallacy (Bragger et al. 1998; 2003).

**5.3.3. Framing.** *Framing* effects occur when people give different answers to structurally similar problems depending on their wording – their “frame” (Tversky & Kahneman 1981). Our intuitions are generally blamed for these effects (Kahneman 2003). Another explanation that can be seen as either complementary or alternative to this one is that different frames make some reasons more or less available, thus modifying the way reasoning affects our decisions. Several results support this interpretation (see McKenzie 2004; McKenzie & Nelson 2003). First, as mentioned earlier, participants who reason more about the tasks are more influenced by framing effects (Igou & Bless 2007). Second, when groups make decisions on framed problems, the groups tend to converge on the answer that is supported by the strongest reasons (McGuire et al. 1987; Milch et al. 2009; Paese et al. 1993). If the participants’ answers were truly based on their intuitions, the answer proposed by the group would tend to be the mean of these different intuitions (Allport 1924; Farnsworth & Behner 1931). Instead, these findings have to be explained within the framework of the Persuasive Argument Theory (Vinokur 1971; Vinokur & Burnstein 1978), showing that the decisions are based on reasons.

**5.3.4. Preference inversion.** The ability to evaluate preferences correctly is necessary for economic models of decision making, but preferences can vary dramatically depending on the way they are measured. Someone may rate A higher than B and still choose B over A (Bazerman et al. 1992; Irwin et al. 1993; Kahneman & Ritov 1994; Slovic 1975; Tversky et al. 1988). For example, the relative rating of two objects can vary or even be reversed,

depending on whether they are rated separately or jointly (Hsee 1996b; 1998; Hsee et al. 1999). Thus, when the following two objects are presented in isolation – a music dictionary with 10,000 entries that is “like new,” and one with 20,000 entries and a torn cover – people rate the one with 10,000 entries more highly. However, when people have to choose between the two, they favor the one that has more entries, despite the torn cover (Hsee 1996b). Such effects fit perfectly in the current framework: People choose an alternative because they can provide “a compelling argument for choice that can be used to justify the decision to oneself as well as to others” (Tversky et al. 1988, p. 372). In the foregoing example, people lack reliable intuitions – they cannot tell how many entries a good music dictionary should have. Lacking such intuitions, they fall back on reasoning and let their judgments be guided by ease of justification – in this case, the condition of the dictionary that easily justifies a high or low price. On the other hand, dimensions with numerical values will often provide compelling justifications when options are presented jointly. This bias can lead to suboptimal decisions (Hsee & Zhang 2004).

More generally, “decision-makers have a tendency to resist affective influence, and to rely on rationalistic attributes to make their decisions” (Hsee et al. 2003, p. 16; see also Okada 2005). Indeed, rationalistic attributes make for easy justifications. For instance, in one experiment, participants had either to choose between the following two options or to rate them: A roach-shaped chocolate weighing 2 ounces and worth 2 dollars, and a heart-shaped chocolate weighing half an ounce and worth 50 cents (Hsee 1999). A majority (68%) of participants chose the roach-shaped chocolate, even though more than half (54%) thought they would enjoy the other more. The participants who chose the bigger, roach-shaped chocolate did so because the feeling of disgust, being “irrational,” was hard to justify, especially compared with the difference in price and size. However, in the light of the results from the psychology of disgust (e.g., Rozin et al. 1986), we can tell that their choice was certainly the wrong one.

**5.3.5. Other inappropriate uses of reasons.** Many other inappropriate uses of reasons have been empirically demonstrated. Investors’ decisions are guided by reasons that seem good but are unrelated to real performance (Barber et al. 2003). People will use a rule such as “more variety is better” or “don’t pick the same things as others” to guide their decisions, even when less variety or more conformity would actually be more in line with their preferences (Ariely & Levav 2000; Berger & Heath 2007; Simonson 1990). Use of a rule such as “don’t pay for delays” will lead to behaviors that go against one’s own interest (Amir & Ariely 2003). When forecasting their affective states, people rely on explicit lay theories (Igou 2004), which will often lead them astray (Hsee & Hastie 2006). Because “it’s better to keep options open,” people will be reluctant to make an unalterable decision even when they would be better off making it (Gilbert & Ebert 2002). When indulging in a hedonic act, people feel they need a reason for such indulgence, even though this does not actually change the quality of the experience (Xu & Schwarz 2009). Reason-based choice has also been used to explain effects related to loss aversion (Simonson & Nowlis 2000), the effect of

attribute balance (Chernev 2005), the tendency to be overwhelmed by too much choice (Scheibehenne et al. 2009; Sela et al. 2009), the feature creep effect (Thompson et al. 2005a), the endowment effect (Johnson et al. 2007), aspects of time discounting (Weber et al. 2007), and several other departures from the norms of rationality (Shafir et al. 1993).

Another sign that reason-based choice can lead to non-normative outcomes is that sometimes reasons that are not relevant to the decision will nonetheless play a role. For instance, the same irrelevant attribute will sometimes be used as a reason for choosing an item (Carpenter et al. 1994) and sometimes as a reason for rejecting it (Simonson et al. 1993; 1994), depending on what decision it makes easier to justify (Brown & Carpenter 2000). People will also be influenced by irrelevant pieces of information because they find it hard to justify ignoring them (Tetlock & Boettger 1989; Tetlock et al. 1996).

All of these experiments demonstrate cognitively unsound uses of reasoning. There are two ways to explain these findings. One could argue that these are instances of a mechanism designed for individual cognition, and in particular for decision making, that sometimes gets misused. According to the argumentative theory, however, the function of reasoning is primarily social: In particular, it allows people to anticipate the need to justify their decisions to others. This predicts that the use of reasoning in decision making should increase the more likely one is to have to justify oneself. This prediction has been borne out by experiments showing that people will rely more on reasons when they know that their decisions will later be made public (Thompson & Norton 2008) or when they are giving advice (in which case one has to be able to justify oneself [see Kray & Gonzalez 1999]). By contrast, when they are choosing for others rather than for themselves, they are less prone to these effects because there is then less need for a utilitarian, justifiable decision (Hamilton & Thompson 2007). Finally, it should be stressed that the picture of reasoning painted in these studies may be overly bleak: Demonstrations that reasoning leads to errors are much more publishable than reports of its successes (Christensen-Szalanski & Beach 1984). Indeed, in most cases, reasoning is likely to drive us towards good decisions. This, we would suggest, is mostly because better decisions tend to be easier to justify. The reasons we use to justify our decisions have often been transmitted culturally and are likely to point in the right direction – as when people justify their avoidance of sunk-cost mistakes by using the rule they have learned in class (Simonson & Nye 1992). In such cases, the predictions of the argumentative theory coincide with those of more classical theories. However, what the results just reviewed show is that, when a more easily justifiable decision is *not* a good one, reasoning still drives us in the direction of ease of justification. Even if they are rare, such cases are crucial to comparing the present theory (reasoning drives us to justifiable decisions) with more classical ones (reasoning drives us to good decisions).

## 6. Conclusion: Reasoning and rationality

Reasoning contributes to the effectiveness and reliability of communication by enabling communicators to argue



for their claim and by enabling addressees to assess these arguments. It thus increases both in quantity and in epistemic quality the information humans are able to share.

We view the evolution of reasoning as linked to that of human communication. Reasoning, we have argued, enables communicators to produce arguments to convince addressees who would not accept what they say on trust; it enables addressees to evaluate the soundness of these arguments and to accept valuable information that they would be suspicious of otherwise. Thus, thanks to reasoning, human communication is made more reliable and more potent. From the hypothesis that the main function of reasoning is argumentative, we derived a number of predictions that, we tried to show, are confirmed by existing evidence. True, most of these predictions can be derived from other theories. We would argue, however, that the argumentative hypothesis provides a more principled explanation of the empirical evidence (in the case of the confirmation bias, for instance). In our discussion of motivated reasoning and of reason-based choice, not only did we converge in our prediction with existing theories, but we also extensively borrowed from them. Even in these cases, however, we would argue that our approach has the distinctive advantage of providing clear answers to the why-questions: Why do humans have a confirmation bias? Why do they engage in motivated reasoning? Why do they base their decisions on the availability of justificatory reasons? Moreover, the argumentative theory of reasoning offers a unique integrative perspective: It explains wide swaths of the psychological literature within a single overarching framework.

Some of the evidence reviewed here shows not only that reasoning falls short of delivering rational beliefs and rational decisions reliably, but also that, in a variety of cases, it may even be detrimental to rationality. Reasoning can lead to poor outcomes not because humans are bad at it but because they systematically look for arguments to justify their beliefs or their actions. The argumentative theory, however, puts such well-known demonstrations of “irrationality” in a novel perspective. Human reasoning is not a profoundly flawed general mechanism; it is a remarkably efficient specialized device adapted to a certain type of social and cognitive interaction at which it excels.

Even from a strictly epistemic perspective, the argumentative theory of reasoning does not paint a wholly disheartening picture. It maintains that there is an asymmetry between the *production* of arguments, which involves an intrinsic bias in favor of the opinions or decisions of the arguer whether they are sound or not, and the *evaluation* of arguments, which aims at distinguishing good arguments from bad ones and hence genuine information from misinformation. This asymmetry is often obscured in a debate situation (or in a situation where a debate is anticipated). People who have an opinion to defend don't really evaluate the arguments of their interlocutors in a search for genuine information but rather consider them from the start as counterarguments to be rebutted. Still, as shown by the evidence reviewed in section 2, people are good at assessing arguments and are quite able to do so in an unbiased way, provided they have no particular axe to grind. In group reasoning experiments where participants share an interest in discovering the

right answer, it has been shown that *truth wins* (Laughlin & Ellis 1986; Moshman & Geil 1998). While participants in collective experimental tasks typically produce arguments in favor of a variety of hypotheses, most or even all of which are false, they concur in recognizing sound arguments. Since these tasks have a demonstrably valid solution, truth does indeed win. If we generalize to problems that do not have a provable solution, we should at least expect good arguments to win, even if this is not always sufficient for truth to win (and, in section 2, we have reviewed evidence that this is indeed the case). This may sound trivial, but it is not. It demonstrates that, contrary to common bleak assessments of human reasoning abilities, people are quite capable of reasoning in an unbiased manner, at least when they are evaluating arguments rather than producing them, and when they are after the truth rather than trying to win a debate.

Couldn't the same type of situation that favors sound evaluation favor comparable soundness in the production of arguments? Note, first, that situations where a shared interest in truth leads participants in a group task to evaluate arguments correctly are not enough to make them produce correct arguments. In these group tasks, individual participants come up with and propose to the group the same inappropriate answers that they come up with in individual testing. The group success is due to, first and foremost, the filtering of a variety of solutions, achieved through evaluation. When different answers are initially proposed and all of them are incorrect, then all of them are likely to be rejected, and wholly or partly new hypotheses are likely to be proposed and filtered in turn, thus explaining how groups may do better than any of their individual members.

Individuals thinking on their own without benefiting from the input of others can assess only their own hypotheses, but in doing so, they are both judge and party, or rather judge and advocate, and this is not an optimal stance for pursuing the truth. Wouldn't it be possible, in principle, for an individual to decide to generate a variety of hypotheses in answer to some question and then evaluate them one by one, on the model of Sherlock Holmes? What makes Holmes such a fascinating character is precisely his preternatural turn of mind operating in a world rigged by Conan Doyle, where what should be inductive problems in fact have deductive solutions. More realistically, individuals may develop some limited ability to distance themselves from their own opinion, to consider alternatives and thereby become more objective. Presumably this is what the 10% or so of people who pass the standard Wason selection task do. But this is an acquired skill and involves exercising some imperfect control over a natural disposition that spontaneously pulls in a different direction.

Here, one might be tempted to point out that, after all, reasoning is responsible for some of the greatest achievements of human thought in the epistemic and moral domains. This is undeniably true, but the achievements involved are all collective and result from interactions over many generations (on the importance of social interactions for creativity, including scientific creativity, see Csikszentmihalyi & Sawyer 1995; Dunbar 1997; John-Steiner 2000; Okada & Simon 1997). The whole scientific enterprise has always been structured around groups, from the Lincean Academy down to the Large Hadron

Collider. In the moral domain, moral achievements such as the abolition of slavery are the outcome of intense public arguments. We have pointed out that, in group settings, reasoning biases can become a positive force and contribute to a kind of division of cognitive labor. Still, to excel in such groups, it may be necessary to anticipate how one's own arguments might be evaluated by others and to adjust these arguments accordingly. Showing one's ability to anticipate objections may be a valuable culturally acquired skill, as in medieval *disputationes* (see Novaes 2005). By anticipating objections, one may even be able to recognize flaws in one's own hypotheses and go on to revise them. We have suggested that this depends on a painstakingly acquired ability to exert some limited control over one's own biases. Even among scientists, this ability may be uncommon, but those who have it may have a great influence on the development of scientific ideas. It would be a mistake, however, to treat their highly visible, almost freakish, contributions as paradigmatic examples of human reasoning. In most discussions, rather than looking for flaws in our own arguments, it is easier to let the other person find them and only then adjust our arguments, if necessary.

In general, one should be cautious about using the striking accomplishments of reasoning as proof of its overall efficiency, since its failures are often much less visible (see Ormerod 2005; Taleb 2007). Epistemic success may depend to a significant extent on what philosophers have dubbed *epistemic luck* (Pritchard 2005); that is, chance factors that happen to put one on the right track. When one happens to be on the right track and "more right" than one could initially have guessed, some of the distorting effects of motivated reasoning and polarization may turn into blessings. For instance, motivated reasoning may have pushed Darwin to focus obsessively on the idea of natural selection and explore all possible supporting arguments and consequences. But, for one Darwin, how many Paleys?

To conclude, we note that the argumentative theory of reasoning should be congenial to those of us who enjoy spending endless hours debating ideas – but this, of course, is not an argument for (or against) the theory.

#### ACKNOWLEDGMENTS

We are grateful to Paul Bloom, Ruth Byrne, Peter Carruthers, Nick Chater, Jon Haidt, Ira Noveck, Guy Politzer, Jean-Baptiste Van der Henst, Deirdre Wilson, and four anonymous reviewers for useful suggestions and criticisms on earlier versions of this article. Our work has been supported by a Ph.D grant of the DGA (Paris) to Hugo Mercier and by the CSMN (Oslo).

#### NOTES

1. Recently, *reasoning* has been used simply as a synonym of *inference* and is then unproblematically attributed to infants (Spelke & Kinzler 2007) or to nonhuman animals (Blaisdell et al. 2006). In this article, however, we use "reasoning" in its more common and narrower sense. The content of the article should make it clear why we see this as a principled terminological choice.

2. Our functional hypothesis will be tested without reference to specific mechanisms (as is common in evolutionary biology). Even if one can ask to what extent attributing an argumentative function to reasoning suggests or favors a specific algorithmic

account, this will not be the focus of this article. There is, in any case, no obvious clash between our functional account and various algorithmic accounts that have been offered, for instance, by Evans (2007), Johnson-Laird (2006), or Rips (1994).

3. In the psychology of reasoning, some tasks can be described as *production tasks* because participants have to produce a logically valid conclusion from a set of premises. However, these tasks are very different from the production of arguments in a debate. In a dialogic context, one starts from the conclusion and tries to find premises that will convince one's interlocutor. It is this meaning of *production* that is relevant here.

4. It should be noted that this spotty record may be partly explained by very artificial conditions: In the vast majority of group experiments, participants are asked to interact with people they don't know and will never meet again, and to perform tasks that have no bearing on their lives outside the laboratory. When any of these factors is made more natural, performance improves. Debates about political matters between laypeople often lead to epistemic improvement (Landemore, in press; Mercier & Landemore, in press). Groups that are used to working together are much more efficient (Michaelsen et al. 1989). And collaborative learning is hugely successful in schools (Slavin 1995).

5. Other, slightly weaker results are obtained for inductive tasks (Laughlin et al. 1991; 2002; 2003; 2006). Debates are also a well-known way of improving comprehension in many domains (e.g., see Anderson et al. 1996; 2001; Foot et al. 1994; Howe 1990; Johnson & Johnson 2007; 2009; Lao & Kuhn 2002; Nussbaum 2008; Nussbaum & Sinatra 2003; Slavin 1995; Smith et al. 2009; Tolmie et al. 1993; van Boxtel et al. 2000; Webb & Palinscar 1996).

6. Incidentally, another advantage of the theory suggested here is that it makes testable predictions about the contexts that should motivate the use of reasoning; namely, contexts in which real or anticipated argumentation takes place. This contrasts with standard dual-process theories, which do not have a principled and testable way of predicting when system 2 reasoning should be triggered.

7. It may be worth mentioning that what general motivation fails to bring about is efficient or unbiased reasoning rather than reasoning per se. If you pay people to get the right answer in, say, the Wason selection task, they may reason more but will still be as biased, and their answer will still be wrong.

8. The Delphi technique is a method of forecasting that can be seen as trying to make the best of the confirmation bias by having different experts critique one another's predictions and justify their own predictions. Its effectiveness shows that, in an appropriate context, the confirmation bias can be conducive to very good performance (Green et al. 2007; Keeney et al. 2001; Powell 2003; Rowe & Wright 1999; Tichy 2004).

9. Note that *motivated*, or *motivation*, as used here do not refer to conscious motivation based on reasons, as in "I'm going to think of arguments supporting this opinion of mine in case someone questions me later." Instead, it refers to processes that influence either the direction or the triggering of reasoning in a mostly unconscious manner. Even though a lawyer, for instance, can consciously trigger reasoning and influence its direction, this is the exception and not the rule. Generally, people (including lawyers) have limited control over the triggering of reasoning or the direction it takes.

10. Attitude polarization is most likely to occur in individuals who hold a very strong attitude with a high degree of confidence. The problem is, then, that these individuals will tend to fall at one end of the attitude scale before reading the arguments, which makes it close to impossible to detect any movement towards a more extreme attitude. This can explain, at least in part, the failed replications of Kuhn and Lao (1996) and Miller et al. (1993).

11. Incidentally, this does not explain all forms of belief perseverance: Other mechanisms may be involved in some instances

(e.g., see Anderson et al. 1980), but the availability of arguments supporting the discredited belief may still be crucial (see Anderson et al. 1985).

12. It has been shown that pigeons fall prey to the fallacy but *only* when no indication was given that they were in such a situation (Navarro & Fantino 2005). The instructions received by human participants always make this point clear, so these experiments confirm the point made by Arkes and Ayton (1999).

## Open Peer Commentary

### Arguing, reasoning, and the interpersonal (cultural) functions of human consciousness

doi:10.1017/S0140525X10002785

Roy F. Baumeister,<sup>a</sup> E. J. Masicampo,<sup>b</sup> and C. Nathan DeWall<sup>c</sup>

<sup>a</sup>Department of Psychology, Florida State University, Tallahassee, FL 32306-4301; <sup>b</sup>Department of Psychology, Tufts University, Medford, MA 02155; <sup>c</sup>Department of Psychology, University of Kentucky, Lexington, KY 40506-0044.

baumeister@psy.fsu.edu.

ej.masicampo@tufts.edu    cnathandewall@gmail.com

<http://www.psy.fsu.edu/~baumeister/index.html>

<http://ase.tufts.edu/psychology/ambady/ej.html>

<http://www.uky.edu/~njdewa2/home.html>

**Abstract:** Our recent work suggests that (1) the purpose of human conscious thought is participation in social and cultural groups, and (2) logical reasoning depends on conscious thought. These mesh well with the argument theory of reasoning. In broader context, the distinctively human traits are adaptations for culture and inner processes serve interpersonal functions.

A long tradition has regarded human thinking as a solitary, if not solipsistic, exercise aimed at facilitating behavior. This has privileged the assumption that reasoning is mainly for enabling individuals to seek the truth. Mercier and Sperber (M&S) have instead invoked an interpersonal dimension: Reasoning is for arguing.

The idea M&S advance dovetails nicely with our own work, which has endorsed the view that uniquely human forms of cognition serve interpersonal functions. One such function is the use and accumulation of accurate knowledge in culture. To be sure, to say that reasoning is for arguing does not mean reasoning is irrelevant to seeking the truth, but people seek the truth collectively, not individually. Humans are cultural animals, which means they use cultural systems as their strategy for improving survival and reproduction (e.g., Baumeister 2005). Hence the distinctively human traits, such as the capacity for reason, are mainly for creating culture, sustaining it, and participating in it. Culture accumulates knowledge across time, and huge swathes of knowledge – from cooking and farming to mathematics, science, and technology – can be mastered only by having many individuals build on one another's advances across generations. Arguing is a vital means by which a cultural group builds its stock of knowledge. Even scientists, in principle the most inveterate truth seekers, have been known to argue, and indeed much of the process of science is conducted as arguing with and about evidence. Individuals who are bred to argue can thus combine to form groups that collectively accumulate increasingly accurate knowledge. Meanwhile, hominids born

with less capacity to argue would fail to participate fully in culture, which may have reduced their ability to survive and reproduce.

The notion that reasoning is for arguing fits nicely with another argument we have endorsed, which is that human thought is for sharing one's thoughts and experiences with others. For more than a century, psychology has regarded William James's famous conclusion that thinking is for doing as an unassailable truism. Yet our own research has led us to entertain a rival hypothesis, that much of thinking is for talking (see Baumeister & Masicampo 2010). This applies particularly to *conscious thought*, defined as the advanced human form of cognition that differs from what most animals have.

Many investigators operationally define conscious thought as those thoughts the person can report to others. Few, however, seem to have heeded the implication that the purpose of conscious thought is precisely for enabling people to tell their thoughts to one another.

The interpersonal bases of thinking are an exciting advance and represent potentially a fundamental change in how the field understands the goals and purposes of human thought, especially conscious thought. There have been two overlapping debates about consciousness in recent decades. One is whether conscious thoughts have any causal influence on behavior. A recent survey suggests a positive answer (Baumeister et al. 2011). The other, more difficult question is what value is served by having thoughts be conscious. That is, could not those same thoughts influence behavior just as effectively without being conscious? It is difficult to make an evolutionary or functional case for advantages of having thoughts inside the individual mind be conscious. But it is easy to make the case for the advantages of being able to communicate thoughts with an organized group, which is what conscious thought enables.

Merely sharing thoughts is already helpful in terms of collective use of information, but M&S's focus on arguing is a huge boost and extension to this line of thinking. We wish we had thought of it! Conscious thought enables people to talk to others and thereby enables small groups to resolve differences. By talking, people can resolve conflicts, influence one another, converge on the truth (aided vitally by reasoning when there are differences), and thereby function together more harmoniously and effectively than nonhuman groups. Nonhuman animals, in contrast, have a relatively simple and unsophisticated language apparatus compared with humans. They resolve conflicts through aggression instead of reconciliation, dominate others instead of attempting to persuade one another, and rely on what appears true in the present environment instead of using logical reasoning to understand that things are not always as they initially seem. Thus, M&S's theory sheds light on what has made human beings such a successful species in using culture as an advanced form of social life to improve survival and reproduction.

Consistent with that emphasis and with M&S's article, we have found that logical reasoning depends on conscious thought (DeWall et al. 2008). These experiments showed that reasoning improved as conscious processing was engaged, and reasoning quickly deteriorated when conscious attention was preoccupied elsewhere. In contrast, logical reasoning performance was unaffected by either preoccupying or engaging (priming) the unconscious processing system.

In short, we view M&S's article as an important complement to other work emphasizing human consciousness as serving interpersonal and cultural functions. Conscious thought enables both reasoning and advanced forms of communication, including arguing. Human evolution was partly based on developing mental powers to enable group members to communicate information for group action, which is the foundation of culture. Conscious logical reasoning and interpersonal arguing are vital parts of that process and thus help to define human nature.

## Regret and justification as a link from argumentation to consequentialism

doi:10.1017/S0140525X10002852

Terry Connolly<sup>a</sup> and Jochen Reb<sup>b</sup>

<sup>a</sup>Eller College, University of Arizona, Tucson, AZ 85721; <sup>b</sup>Singapore Management University, Singapore 178899.

Connolly@u.arizona.edu jochenreb@smu.edu.sg

<http://management.eller.arizona.edu/faculty/tconnolly.asp>

[http://www.business.smu.edu.sg/faculty/organisational\\_behavior/jreb.asp](http://www.business.smu.edu.sg/faculty/organisational_behavior/jreb.asp)

**Abstract:** Mercier and Sperber (M&S) argue that reasoning has evolved primarily as an adjunct to persuasive communication rather than as a basis for consequential choice. Recent research on decision-related regret suggests that regret aversion and concomitant needs for justification may underpin a complementary mechanism that can, if appropriately deployed, convert M&S's facile arguer into an effective decision maker, with obvious evolutionary advantages.

Mercier and Sperber (M&S) make the provocative case that, in evolutionary terms, reasoning is better seen as an adjunct to communication than as a guide to decision making. However, since there are also evolutionary advantages to effective consequential choice, broadly interpreted, what might this ability be based on? We argue that emotional responses, specifically those associated with regret aversion and justification, may serve such a role, linking argument making of the sort described by M&S to consequential decision making.

In a continuing program of research, we have shown that regret aversion can help in overcoming decision errors. Much of this research draws on *decision justification theory* (Connolly & Zeelenberg 2002; Connolly et al. 1997), which distinguishes regret associated with a (comparatively) poor outcome (*outcome regret*) from that associated with the judgment that the focal decision was wrong or poorly made – that is, was “unjustified” (*self-blame* or *process regret*). Efforts to avoid regret of this latter sort facilitates improved decision processes (Reb & Connolly 2010), information search (Reb 2008) and task learning (Reb & Connolly 2009).

It also appears to reduce or eliminate reason-based decision errors, such as those discussed in M&S sections 5.2 and 5.3. For example, Connolly et al. (2010) compared the effects of external accountability and regret priming on the attraction (or *decoy*) effect, in which an option is seen as more desirable when it dominates an irrelevant decoy option. Replicating earlier studies (Simonson & Nye 1992; Slaughter et al. 2006), we showed that accountability (a demand to justify one's choice to others) exacerbated the attraction effect, consistent with M&S's argument. Regret priming, in contrast, with its demand to justify one's decision to oneself, eliminated the effect. It seems that making regret salient may have led to a more balanced use of reasoning whose goal was less to convince others and more to arrive at a choice that satisfies one's own values and standards.

Reb (2005) showed that regret priming also reduced or eliminated other “reason-based” effects such as the compromise effect (Simonson 1989), in which an option is more desirable when presented as a compromise, and the accept/reject effect (Shafir & Tversky 1992), in which the same option tends to be both rejected and selected. In all these reason-based choice effects, the justifying arguments do not withstand close scrutiny. They are simply “shallow but nice-sounding rationales” (Simonson 1989, p. 170) that might serve to convince an uncritical external audience but not one's thoughtful self. In contrast, regret priming did not reduce the most important attribute effect (Slovic 1975) where the justifying argument can reasonably be construed to both self and others as a legitimate tiebreaker between equally valued options (Reb 2005).

Regret priming appears to involve both motivational and attention-directing effects, which are sometimes quite subtle. For example, Reb and Connolly (2009) used unobtrusive priming of

either outcome or self-blame regret in a repeated decision task where feedback on outcomes of unchosen options was offered. Subjects primed for outcome regret rejected such feedback more often, learned more slowly, and ultimately performed less well than those primed for self-blame regret (thus falling victim to the *myopic regret aversion* trap: avoiding short-term regret led them to experience greater long-term regret). Both groups were motivated to avoid regret, but one did so by avoiding painful comparisons, the other by following a justifiable decision process.

In summary we find persuasive M&S's case that reasoning is primarily for persuasive argumentation rather than for effective consequential choice. Given the evolutionary advantages of the latter, however, it is plausible that other systems may have developed to support such choice processes. A growing body of evidence suggests that mechanisms of regret, regret avoidance, and justification can serve such a decision-improving role. Specifically, aversion of process regret may complement the fluent argument maker and tweak it to pay more balanced attention to and weighing of the pros and cons associated with a decision problem. Because of the anticipatory nature of regret, attention may further be directed to future consequences that are predicted to impact experienced regret. Mechanisms of regret and justification thus suggest important linkages between the argument-rich communicator sketched by M&S and the purposive consequentialist demanded by rational choice models of human decisions. We see such evidence as dovetailing neatly with, and modestly extending, the findings compiled in the target article. Perhaps ironically, as the central role of reasoning in assuring good choices has come increasingly into doubt in recent decision research, emotions, earlier seen as an obstacle to effective decision making, are increasingly being found to perform crucial functions in facilitating such decisions.

## The freak in all of us: Logical truth seeking without argumentation

doi:10.1017/S0140525X10002827

Wim De Neys

Centre National de la Recherche Scientifique (CNRS)—Université de Toulouse, Maison de la Recherche, 31058 Toulouse Cedex 9, France.

Wim.deneys@univ-tlse2.fr <http://www.univ-tlse2.fr/lrc/deneys>

**Abstract:** Mercier and Sperber (M&S) sketch a bleak picture of logical reasoning in classic, nonargumentative tasks. I argue that recent processing data indicate that despite people's poor performance they at least seek to adhere to traditional logical norms in these tasks. This implies that classic reasoning tasks are less artificial – and logical reasoning less exceptional – than M&S's framework suggests.

Mercier and Sperber (M&S) argue that the notoriously bad logical reasoning performance in classic reasoning and decision-making tasks can be attributed to the lack of argumentative context or interaction in these tasks. They point out that when the same problems are put in an argumentative context, people have little trouble solving them. From this they conclude that, except for a few “almost freakish” (sect. 6, para. 7) individuals, people will engage in a genuine logical reasoning process only when arguing. Clearly, this seems to question the validity of classic reasoning tasks: In these nonargumentative tasks, people will typically not do what they do in the real (argumentative) world. This impression is further strengthened by M&S's claim that it would be a mistake to treat as paradigmatic examples of human reasoning those few individuals who do exert control over their biases and manage to solve the classic reasoning tasks.

I want to point out that although M&S nicely demonstrate that providing an argumentative context can boost people's logical reasoning performance, this does not imply that people do not

already seek to reason logically in the absence of this context. It should be stressed that M&S's claims with respect to the poor performance in classic reasoning tasks are typically based on traditional studies that focused merely on accuracy data (i.e., the output of the reasoning process).

Recent studies that examined a wider range of processing measures such as latency or brain-activation data (e.g., Bonner & Newell 2010; De Neys et al. 2008) sketch a more optimistic picture of people's reasoning performance in the classic, nonargumentative tasks. These data suggest that although people very often fail to select the correct logical response, they at least seek to adhere to the logical norm. For example, although people typically fail to solve classic reasoning problems in which intuitive beliefs conflict with normative considerations, latency studies indicate that people do take longer to respond to these problems compared to problems in which the norms are not being violated (e.g., Bonner & Newell 2010; De Neys & Glumicic 2008). Problems in which cued intuitive beliefs conflict with logical considerations are also longer inspected and better recalled (e.g., Ball et al. 2006; De Neys & Glumicic 2008).

Neuroscientific research further suggests that brain areas involved in the detection of conflict between competing responses are activated when people err and violate a logical norm (e.g., De Neys et al. 2008; 2010). Clearly, if people were not at least engaged in some elementary logical processing and tried to adhere to the logical norm, it is hard to see why violating it would affect their inference process. In addition, De Neys and Franssens (2009) observed that after solving problems in which the intuitive believability and logical validity of a conclusion conflicted, reasoners showed an impaired access to words that were associated with the intuitive beliefs. Such an impaired memory access is considered a key marker of inhibitory processing. Even people who were always biased by their beliefs showed a minimal impairment, indicating that they had attempted to inhibit the intuitive beliefs but failed to complete the process. Once again, if people were not trying to adhere to the logical norm, there would be no reason for them to block the conflicting belief-based response.

The crucial point is that these studies suggest that even without an argumentative context people are already engaging in a logical reasoning process. What the "freakish" individuals who give the correct response seem to be better at is completing the inhibition of conflicting intuitive heuristic responses (De Neys & Franssens 2009; Houdé et al. 2000). However, the important finding in the studies cited is that all reasoners are at least engaging in this inhibition process and try to adhere to the logical norm. In that sense we're all freaks who seek logical truth when solving classic reasoning tasks. The bottom line is that this indicates that the standard tasks are less artificial – and logical reasoning in these tasks less exceptional – than M&S's framework might suggest.

In sum, M&S convincingly demonstrate that human reasoning can benefit from an argumentative context. By pointing to recent processing data, I tried to clarify that this does not necessarily imply that people simply fail to engage in a logical reasoning process in the absence of such a context. This should give pause for thought before drawing strong negative conclusions with respect to the validity of classic reasoning tasks or the illogical nature of people's reasoning in these tasks.

## Reasoning as a lie detection device

doi:10.1017/S0140525X10002815

Jean-Louis Dessalles

Telecom ParisTech, F-75013 Paris, France.

jl@dessalles.fr <http://www.dessalles.fr>

**Abstract:** The biological function of human reasoning abilities cannot be to improve shared knowledge. This is at best a side effect. A more

plausible function of argumentation, and thus of reasoning, is to advertise one's ability to detect lies and errors. Such selfish behavior is closer to what we should expect from a naturally selected competence.

I fully support the central claim by Mercier & Sperber's (M&S) that deliberative reasoning is a by-product of argumentative competence. But if the function of reasoning is argumentation, what is the (biological) function of argumentation? According to (M&S), argumentative reasoning improves "both in quantity and in epistemic quality the information humans are able to share" (sect. 1.2, para. 9) and, thanks to it, "human communication is made more reliable and more potent" (sect. 6, para. 2).

If the biological function of reasoning is to achieve *shared knowledge optimization* (SKO), as suggested in the target article, then why do people show obvious limitations such as confirmation bias? M&S answer that information quality is optimized, not at the individual level, but at the group level. It would even be a good thing that individuals specialize on their (probably erroneous) line of reasoning, as long as argument exchange restores global information quality. The problem is that natural selection does not operate at the collective level. Shared knowledge belongs to the phenotype of no one.

How does the speaker benefit from uttering an argument? If the purpose is to correct or update her own earlier beliefs, why go public with it? And if it is to correct or update others' beliefs, what's her advantage? M&S's explanation for the existence of deliberative reasoning does not escape the general evolutionary paradox of communication: If it benefits listeners only, there should be no speakers; and if it benefits speakers only (for example, by allowing manipulation), there should be no listeners. Invoking collective benefits does not offer an escape route if we wish to remain on firm Darwinian ground.

To solve the paradox, we must depart from SKO. My proposal (Dessalles 1998) is that humanlike reasoning started with logical *consistency checking* (CC), and that humans used it as a *lie detection* (LD) device. As a response to the risk of appearing self-contradicting, the ability to *restore consistency* (RC) through argumentation emerged. In this game, information quality is not what is at stake. The point for individuals is to *advertise* (AD) their ability to perform or resist LD. This advertisement behavior makes sense within a costly signaling model of human communication (Dessalles 2007; 2008).

The main difference with M&S's position comes from AD. M&S are close to the CC/RC distinction when they speak of *evaluation* vs. *production* (of arguments). They fail, however, to see that these two faculties did not evolve for the sake of any form of public knowledge, but as *signals*. Individuals who can publicly signal lies or errors by naming inconsistencies (CC) derive immediate social benefit (Dessalles 2007). Those who publicly restore consistency (RC) gain social benefit, as well, or regain their momentarily lost status.

Contrary to SKO, the competitive nature of AD explains why reasoning is far from remaining a private activity: Argumentation takes up the major part of the 16,000 words spoken daily, on average (Mehl et al. 2007). Moreover, various observations by M&S make more sense within AD rather than SKO, especially the fact that people are better at finding inconsistencies in others' line of reasoning and at finding support for their own. Another argument in favor of AD is the futility of many conversational topics, which makes no sense from an SKO perspective. Yet another good example of the divergence between AD and SKO is offered by the *BBS* commentary system: Commentators are of course concerned by the overall quality of scientific knowledge, but most of them are even more motivated by the urge to show their ability to point to some inconsistency in the target article. SKO would perhaps hold if contributors accepted that their names be omitted.

M&S strangely do not mention a fundamental common property between deliberative reasoning and argumentation. Both processes seem to consist in a *sequential* alternation between logical incompatibilities and attempts to resolve them. This

property is concisely captured by the conflict-abduction-negation procedure that describes argumentative processes (Dessalles 2008). The sequential nature of argumentative reasoning supports the central claim of the target article, but it is at odds with any idea of knowledge optimization. Virtually all artificial reasoning devices (from chess players to planning programs) involve parallelism whenever possible (especially in mutation comparison). So-called truth-maintenance systems and argumentation systems make use of graph representations that are *not* limited to sequential processing (e.g., Dung 1995). In comparison, human argumentative reasoning is skewed. It is bound to start from a logical incompatibility, and then sequentially creeps forward through recursive attempts to solve the current incompatibility and then detect new ones. Such manifestly sub-optimal procedure does not make sense if the aim is knowledge optimization. It makes perfect sense, however, in the LD/AD context.

The biological importance of informational capabilities is a consequence of the particular political context of our species (Dessalles 2007). In that context, information is not important as such; it is rather an excuse to show off informational capabilities, such as being the first to point to unexpected events. In the absence of a lie detection system, such communication is bound to checkable, almost immediate, events. The advent of CC capabilities offered a new occasion for individuals to compete, by allowing them to advertise their lie and error detection capabilities. This new competition has side effects, such as opening the possibility of communicating about past events that cannot be checked directly. Knowledge improvement also turns out to be a side effect of reasoning and argumentation. When reasoning and producing arguments, speakers follow a more selfish agenda, which is to show off their competence for dealing with anomalies in information.

## Reasoning is for thinking, not just for arguing

doi:10.1017/S0140525X10002773

Jonathan St. B. T. Evans

Centre for Thinking and Language, School of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom.

j.evans@plymouth.ac.uk

**Abstract:** There is indeed extensive evidence that people perform fairly poorly in reasoning tasks and that they often construct arguments for intuitively cued responses. Mercier & Sperber (M&S) may also be right to claim that reasoning evolved primarily as argumentation. However, if it did, the facility became exapted to the function of supporting uniquely human abilities for reflective thinking and consequential decision making.

A number of claims are made in the target article that are difficult to dispute. People do appear to be skilled at everyday argumentation while struggling to solve reasoning problems presented in the psychological laboratory. There is a great deal of evidence in the psychology of reasoning and decision making – as the authors demonstrate with admirable scholarship – that people frequently use reasoning to justify intuitively cued responses. And of course it is much easier to argue that people evolved skills of argumentation than the ability to solve reasoning problems. However, as is commonly observed by evolutionary theorists, a facility which evolves for one function may become “exapted” for another. Hence, my only concern is that the authors may have used their own formidable skills of argumentation to overstate the case against explicit reasoning as a tool for rational thinking and decision making.

As Mercier and Sperber (M&S) note, their analysis has implications for the now ubiquitous dual-process theories of higher cognition (Evans 2008). It is interesting to note that the very

first such theory in the psychology of reasoning was consistent with the authors’ proposals. Wason and Evans (1975; Evans & Wason 1976) distinguished between type 1 (intuitive) processes, which unconsciously cued the relevance of certain cards on the Wason selection task, and type 2 (explicit reasoning) processes, which served to confabulate justifications for these choices. Although it seems clear now that people do reason on the selection task, there is strong evidence that this is focused on finding justifications for unconscious cued cards. However, where such a justification cannot be found, recent evidence indicates that participants may withhold an intuitively prompted response (Evans & Ball 2010). Hence, even in this case, reasoning is performing some beneficial role in decision making.

M&S are also correct to say that in more recent years dual-process theorists have emphasised the efficacy of type 2 processing in problem solving and decision making, especially for novel problems (e.g., Evans 2010; Evans & Over 1996; Kahneman & Frederick 2002; Stanovich 1999; 2010). In particular, theorists have emphasized the ability of some individuals, especially those of higher cognitive ability, to engage in hypothetical thinking and mental simulation, decoupling their actual beliefs in order to support suppositional reasoning (Evans 2007; Stanovich 2010). Such theorists attribute the unique cognitive and cultural achievements of the human species to the development of a second system of cognition or a *new mind*, largely unavailable to other animals. Moreover, it is proposed that while both new and old minds have instrumental rationality, they pursue different goals by different cognitive mechanisms (Evans 2010; Stanovich 2004). In particular, the old mind (broadly shared with higher animals) is driven by the past, replicating by evolution or individual learning past successes, whereas the new mind (distinctively human) can conduct mental simulations and reason consequentially, anticipating the future.

M&S deal with this issue rather dismissively, stating that “the possibility to deal with novelty and to anticipate the future is less a characterization of reasoning than it is of learning” (sect. 1.2, para. 3), going on to argue that the occasional successes of reasoning in helping us to solve problems would be insufficient to explain its evolution. It is possible to agree with the latter claim while strongly disputing the former. First, the learning mechanisms of the old mind do only enable us to repeat what has worked in the past. For novel problems which require simulation of future events, such a mechanism is largely useless. And while we are admittedly fairly poor at consequential decision making (Baron 1994) we *can* do it. Since 1945, politicians have so far resisted their traditional practice of applying their most powerful weapons to their enemies when such weapons are nuclear, presumably because of the anticipated catastrophic consequences. And while it is debatable whether the human race will in fact avoid the disastrous effects of climate change, it could hardly be denied that the world is currently engaged in detailed hypothetical thinking (aided by computer models) about the future of the planet, and that much current policy is being driven by this. No other animal can even remotely think and act in this way.

M&S, as other evolutionary psychologists before them, assert the domain-specific nature of human reasoning and apparently overlook the importance of heritable general intelligence (for detailed critique of such arguments, see Stanovich 2004; Stanovich & West 2003). Research on dual-process theory has provided a mass of evidence that the ability to solve novel reasoning and decision problems is related to individual differences in general intelligence, working-memory capacity, and other (highly correlated) measures of cognitive ability. Solving novel problems is also related to rational thinking dispositions, which help to determine (Stanovich 2009; 2010) the tendency to engage reflective reasoning rather than rely on intuition (see also Frederick 2005). Such findings fly in the face of fashionable claims (e.g., Dijksterhuis et al. 2006b; Gigerenzer 2007; Gladwell 2005) that we are better off relying on intuition than engaging our powers of reasoning.

The difficulty which such advocates of intuition have faced is to explain why humans evolved a capacity for reasoning which is best not trusted. M&S attempt to fill that gap in the target article, thus supporting what I believe to be a dangerously flawed line of reasoning about human intelligence. It is not necessary to follow them down this path in order to respect the integrity of their arguments about the evolution of reasoning via argumentation. Unique human abilities for reflective thought have required the evolution of a number of facilities, including language, metarepresentation, and large forebrains, none of which could plausibly have been driven by some Darwinian need for a new mind. If there were such a driver, surely other animals would have evolved human-like intelligence. It is more plausible to argue that the new mind was an evolutionary accident, in which case an exapted ability for reasoning derived from argumentation may well be part of that story.

### Artificial cognitive systems: Where does argumentation fit in?

doi:10.1017/S0140525X10002839

John Fox

Department of Engineering Science, University of Oxford, Oxford OX1, United Kingdom.

John.fox@eng.ox.ac.uk www.cossac.org

**Abstract:** Mercier and Sperber (M&S) suggest that human reasoning is reflective and has evolved to support social interaction. Cognitive agents benefit from being able to reflect on their beliefs whether they are acting alone or socially. A formal framework for argumentation that has emerged from research on artificial cognitive systems that parallels M&S's proposals may shed light on mental processes that underpin social interactions.

Mercier and Sperber (M&S) offer a provocative view of argumentation as *reasoning for social purposes*. Human reasoning, they suggest, is not the same as classical *inference* in the sense that in reasoning, the *rationale* for conclusions is available for reflection and hence for communication and discussion. This is an important distinction, but there are also grounds for believing that reflective reasoning has general benefits for any cognitive agent, not just for social interaction.

A domain in which these benefits are evident is reasoning and decision making in medicine. I have a long-standing interest in the cognitive mechanisms that support decision making and other high-level cognitive processes that underpin human expertise, and argumentation has acquired a central role in our work. Early approaches based on logical and probabilistic simulations of cognitive processes yielded promising results (Fox 1980), but extending either model to capture the flexible and adaptive character of human thinking proved difficult. Among the reasons for this were that there was no representation of the rationale on which to reflect – to question prior conclusions or the relevance of evidence, for example.

Subsequent work has sought to address this. This research programme has focused on artificial intelligence (AI) rather than psychology, so my comments should be taken as complementary to the M&S hypothesis rather than directly addressing it. However, I will suggest that a cognitive agent, whether human or artificial, derives major benefits from being able to reflect on its mental states; its goals, intentions, justifications for its beliefs and so on (Das et al. 1997; Fox & Das 2000; Fox et al. 1990). Metacognitive capabilities confer flexibility and robustness whether an agent is acting alone or in concert with others.

Mercier and Sperber's (M&S's) distinction between *inference*, which they call "intuitive," and *reasoning*, which affords "reflection," may perhaps be clarified by a formal perspective. A standard way of formalizing inference systems is to provide a

"signature" that specifies how one set of sentences (e.g., propositions) is entailed by another set of sentences (e.g., a database of propositions and rules). This is a typical inference signature:

$$\frac{\text{Database}}{\text{Conclusion}}L \text{ Inference}$$

That is to say: Conclusion can be validly inferred from Database under the axioms of inference system *L*.

Complex cognitive tasks like decision making and planning require a more complex signature. To emulate human clinical decision making, we sought a reasoning model in which *general medical knowledge* is applied to *specific patient data* by arguing the pros and cons of alternative ways of achieving clinical goals. This is summarized by the following signature.

$$\frac{\text{Knowledge} \cup \text{Data}}{(\text{Claim, Grounds, Qualifier})}LA \text{ Argumentation}$$

In contrast to the *atomic* conclusion of the inference signature, this formulation makes the *structure* of arguments explicit. In LA, a Logic of Argument (Fox et al. 1993), the structure distinguishes three things: the Claim (a tentative conclusion), Grounds (justification), and Qualifier (the confidence in the Claim warranted by the argument. As in classical decision theory, but not classical logic, *collections of arguments* can be aggregated within the LA framework to yield an overall measure of confidence in competing claims. For example, an agent may have multiple lines of argument for and against competing diagnoses or treatments, each of which increases or decreases overall confidence.

LA was developed for cognitive tasks like situation assessment, decision making, and planning, which often involve uncertainty. Uncertainty is modelled explicitly by means of the Qualifier and therefore permits reflection. A qualifier may indicate that an argument "supports" or "opposes" a claim, for example. In *The Uses of Argument* the philosopher Stephen Toulmin has also pointed out that people routinely use *linguistic qualifiers* such as "presumably...", "possibly...", "probably...", and their lexical and affixal negative forms; linguistic qualifiers can be formalised as conditions for accepting claims based on collections of arguments (Elvang-Goransson et al. 1993). Quantitative schemes for expressing argument *strength*, such as Bayesian representations (e.g., Oaksford and Chater [2009] discussion in *BBS* vol. 32) can also be accommodated within the framework (Fox 2003; Fox et al. 1993).

It is a truism that the more supporting (opposing) arguments there are for a claim, the more (less) confidence we should have in it, which we have called the *evidential mode* (Fox, in press). Another mode, *dialectical argumentation*, exploits the observation that discussion and debate also commonly involves "attacks" which rebut or undercut the arguments of other agents. Researchers in AI and computational logic are giving substantial attention to argumentation for modelling interactions and dialogues between cognitive agents (Besnard & Hunter 2008). Argumentation theory may therefore offer insights into the kinds of social interactions that M&S are investigating.

Formal argumentation theory has practical applications. LA is the foundation of *PROforma*, a language for modelling cognitive agents (Fox & Das 2000; Fox et al. 2003); which has been used to develop many practical decision tools, notably in medicine (OpenClinical 2001–6). Argumentation theory may also help to clarify the philosophical and theoretical nature of somewhat vague notions like *evidence*, as this term is commonly used in legal, medical, scientific, and other kinds of reasoning and in everyday decision-making and evidence-based discussions (OpenClinical 2001–6).

These practical uses of argumentation theory do not directly address M&S's proposition that human cognition has evolved to support argument-based reasoning, but the practical power

of argumentation techniques seems to corroborate their belief that there are good evolutionary reasons why humans argue. I do not intend to make psychological claims or predictions based on the specifics of LA, however. The logic was not developed as a descriptive model, and more recent argumentation systems do not all comply with its signature. However the general approach may offer a framework for articulating and comparing psychological theories of inference and reasoning and the cognitive processes that depend on them.

It appears that there are significant convergences between M&S's psychological proposals and developments in AI and computer science which have been driven by interests in artificial cognitive systems. Argumentation is a new and versatile reasoning paradigm that is being studied in many different fields; it may have wide implications for general theories of cognition.

## Reasoning, argumentation, and cognition

doi:10.1017/S0140525X10002979

Keith Frankish

*Department of Philosophy, The Open University, Milton Keynes MK7 6AA, United Kingdom.*

[k.frankish@open.ac.uk](mailto:k.frankish@open.ac.uk)

<http://www.open.ac.uk/Arts/philos/frankish.htm>

**Abstract:** This commentary does three things. First, it offers further support for the view that explicit reasoning evolved for public argumentation. Second, it suggests that promoting effective communication may not be the only, or even the main, function of public argumentation. Third, it argues that the data Mercier and Sperber (M&S) cite are compatible with the view that reasoning has subsequently been co-opted to play a role in individual cognition.

I agree with Mercier and Sperber (M&S) on the need for a dual-process perspective that distinguishes intuitive inference and explicit reasoning, and I find M&S's case for the argumentative theory of reasoning attractive. The theory is also a salutary corrective to the tendency of dual-process theorists to see explicit ("type 2") reasoning as unbiased and normatively correct. (Henceforth, I shall follow M&S in using "reasoning" for explicit thought processes, as opposed to non-conscious intuitive inferences.)

Here I shall add some brief, broadly sympathetic comments. First, I want to outline a further reason for thinking that reasoning evolved for social, argumentative purposes. M&S claim that reasoning is a personal-level activity, involving the construction of complex arguments in response to intuitive beliefs about the strength of the component steps. This view, which I have myself defended in various places (e.g., Frankish 1998; 2004; 2009), requires that reasoners have intuitive-level knowledge of rules of inference, such as *modus ponens*, which guides their explicit reasoning. (*Rules* here includes rules of thumb; the arguments we construct need not be demonstrative; see Frankish 2004.) Now, there are two main forms these rules could take: They could be abstract, defined over concepts and logical structures, or they could be linguistic, defined over the words and structures of a natural language. (Rules of the latter kind can be applied in constructing arguments, provided the arguments are expressed linguistically, either overtly or in inner speech.) They are far more likely to take the latter form, however, especially in early human societies. Knowledge of linguistic rules can be obtained relatively easily in the course of exposure to, and engagement in, argument with one's peers. (It might even be purely procedural, embedded in routines for manipulating linguistic structures.) Knowledge of abstract rules, by contrast, would require explicit teaching, or at least abstraction from previously acquired linguistic rules. (These considerations are set out at more length in Frankish 1998 and 2004.) Note

that I am assuming here that the knowledge that guides explicit argument construction is largely learned. The case for this is strong; reasoning strategies are known to be responsive to verbal instruction and to vary dramatically between individuals (e.g., Stanovich 1999). Note, too, that this view is compatible with M&S's claim that there is a specialized intuitive mechanism for representing arguments. It merely implies that the mechanism operates primarily on linguistic representations and that its rule database is acquired.

If this is right, then it supports the view that the original function of reasoning was social and argumentative. For it suggests that individual reasoning is an internalized version of overt argumentation, conducted in inner speech and guided by knowledge acquired in the course of public argumentation. (There are other reasons, too, for thinking that conscious reasoning is language-involving; see, e.g., Carruthers 1996 and 1998.) And this in turn suggests that public argumentation predated individual reasoning and that whatever adaptations we have for reasoning originally evolved to serve the former. (We can tell a bootstrapping story about the origins of the knowledge that guides public argumentation, involving a process of cultural evolution.)

Second, I want to say something about the function of public argumentation. M&S argue that this is to make communication more efficient and reliable. I agree that argumentation does this, but it may not be its only, or even main, function. As the data M&S cite make clear, group reasoning is often a strikingly effective process of inquiry, which zeroes in on the correct solution. Reasoning may thus have evolved primarily for *collective* cognition, if not for the individual variety. (It would not be surprising if evolution preferred collective cognitive mechanisms. Early human societies would have depended heavily on collective action and problem solving, whereas individual projects might have disrupted group cohesion.) Argumentation may have other important social functions, too. For example, displays of argumentative prowess may serve to attract mates, providing evidence of health and mental acuity.

Third, while I agree that reasoning evolved to serve public argumentation (and still does serve it), I suspect M&S underestimate the extent to which it has subsequently been co-opted to play a role in individual cognition. For the demands of argumentation and individual cognition may not be as different as M&S suppose. There are two points here. First, sound arguments tend to be convincing and optimal decisions defensible, so an argumentative mechanism will often deliver the same results a cognitive mechanism would. Secondly, the biases natural to argumentation may also affect much individual reasoning. For this too may often have a quasi-argumentative structure, with individuals seeking to defend their attitudes and decisions against criticism from their future selves.

I shall expand on the latter point. The idea is that, even as private reasoners, we have other motives besides epistemic ones. We often accept propositions for pragmatic reasons, because we find them comforting, or because they are popular with our peers. Moreover, we value psychological stability for its own sake; we like to have a settled set of opinions (to know our own minds) (Frankish 2004). We may therefore be inclined to bolster our opinions with arguments, so that our future selves will be able to resist counter-evidence and avoid distressing changes of mind. Thus, even in private we may tend to display confirmation bias and motivated, proactive reasoning. Something similar goes for decision making. Intentions serve to coordinate future planning (e.g., Bratman 1987), and in many cases it is better to form and stick with an intention, even if it is not optimal, than to remain undecided or keep changing one's plans. Given this, people may tend to settle on decisions they find easy to justify to themselves, and to buttress them with arguments, so that they will not be tempted to revise them later. Hence, they may still tend to exhibit reason-based choice, even when they do not have an eye on the reactions of their peers.



Thus, while the results cited by M&S may show that reasoning is not well adapted for individual intellectual inquiry (which, as M&S rightly stress, we find very hard), they do not show that it is not adapted to other roles in individual cognition, broadly construed. Of course, as M&S note, motivated reasoning and reason-based choice often have unwelcome consequences (especially, perhaps, in modern technological societies), but, if anything, this tends to support the present suggestion, since the more functions these biases have, the more gains there are to offset the costs.

## Reasoning as deliberative in function but dialogic in structure and origin

doi:10.1017/S0140525X10002906

Peter Godfrey-Smith<sup>a</sup> and Kritika Yegnashankaran<sup>b</sup>

<sup>a</sup>Department of Philosophy, Harvard University, Cambridge, MA 02138;

<sup>b</sup>Department of Philosophy, Bard College, Annandale-on-Hudson, NY 12504.

pgs@fas.harvard.edu kyegnash@bard.edu

<http://www.people.fas.harvard.edu/~pgs/>

**Abstract:** Mercier and Sperber (M&S) claim that the main function of reasoning is to generate support for conclusions derived unconsciously. An alternative account holds that reasoning has a deliberative function even though it is an internalized analogue of public discourse. We sketch this alternative and compare it with M&S's in the light of the empirical phenomena they discuss.

Mercier and Sperber (M&S) argue that the function of reasoning is *argumentative*: "It is to devise and evaluate arguments intended to persuade" (see their abstract). This contrasts with a more familiar *deliberative* view of reasoning, which holds that the function of reasoning is to draw new conclusions and form new beliefs. Reasoning within that more familiar view is then seen as a special kind of *inference*, perhaps one with a distinctive relationship to consciousness and the rational faculties of the whole agent. Such views also tend to be individualistic; they hold that the psychology of reasoning has no special relation to social life.

M&S do allow that sometimes reasoning leads to new conclusions on practical and theoretical matters being drawn by the reasoner, conclusions that can be put to use in guiding action. But this is an incidental by-product of reasoning's main function, where "function" is understood in evolutionary terms.

There is also a third option, however, one drawing on the views of the Russian psychologist Lev Vygotsky (1986). On this view, reasoning is deliberative in function but dialogic in structure (Yegnashankaran 2010). Reasoning is an internalized analogue of interpersonal discourse. Interpersonal discourse itself might be typically a complicated mix of attempts to persuade, attempts to think things through and form new conclusions, and other activities, but what results in our psychology is a tool whose function is primarily deliberative. We do not think that this view is clearly superior to M&S's, but we do think it is an important option to have on the table when considering the evolution of reasoning and the opposition between deliberative and argumentative views.

Once we have the contrast between M&S's view and the Vygotskian version of the deliberative view in mind, the message of the empirical evidence is less clear. M&S say that, on their view, "reasoning should produce its best results when used in argumentative contexts, most notably in group discussions" (sect. 1.2, para. 11). This, they say, is what we actually find. But if the aim of reasoning is to help in persuasion, one would think that a context of dialogue would promote more and more agile deployment of justifications for whatever each agent antecedently believes, not a willingness to respond to others' arguments by changing one's mind. M&S see people as poor individual reasoners but "skilled arguers," where skilled arguers "are not after the truth but after arguments supporting

their views" (see their abstract). But that picture is at tension with the fact that people interacting in groups are, as M&S accept, quite *good* at finding the truth by exchanging ideas, and not merely at buttressing their own positions. And on the M&S view as we understand it, any similarity between changes of mind induced by the social exchange of ideas and changes of mind induced by private reflection is incidental.

On the other side, some forms of confirmation bias do fit better with M&S's view. On a Vygotskian deliberative view, an agent has no good reason to prefer a search for confirmation of a hypothesis they are inclined to believe, to a search for disconfirmation of the hypothesis. On M&S's view, this tendency does make sense.

Finally, we suggest that M&S may underestimate the adaptive value of the directions agents may be led by conscious reasoning. For example, they discuss an experiment where individuals are asked to choose between a small heart-shaped chocolate and a larger chocolate shaped like a roach. Most individuals chose the roach-shaped one, because making the other choice would be harder to rationally justify. M&S say that "in the light of the results from the psychology of disgust . . . , we can tell that their choice was certainly the wrong one" (sect. 5.3.4, para. 2). But if an analogue of this chocolate choice was faced in an evolutionary setting, a reasoner would win out.

## Understanding, evaluating, and producing arguments: Training is necessary for reasoning skills

doi:10.1017/S0140525X1000292X

Maralee Harrell

Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213.

mharrell@cmu.edu

<http://www.hss.cmu.edu/philosophy/faculty-harrell.php>

**Abstract:** This commentary suggests that the general population has much less reasoning skill than is claimed by Mercier & Sperber (M&S). In particular, many studies suggest that the skills of understanding, evaluating, and producing arguments are generally poor in the population of people who have not had specific training.

The target article by Mercier & Sperber (M&S) offers several arguments for their Reasoning is Argumentation hypothesis – that the primary function of reasoning in human beings is to evaluate and produce arguments intended to persuade. While I believe that the Reasoning is Argumentation hypothesis is interesting and should be explored, my comments focus on one specific claim M&S make.

To show that the predictions of their hypothesis are borne out, M&S point to multiple psychological studies that purport to demonstrate that people are generally able to reason well. In this context, reasoning well consists in being able to understand, evaluate, and produce arguments. In particular, M&S claim that studies show that (1) people are good at evaluating both subarguments and overall arguments, and (2) people can generally produce good arguments in a debatelike setting.

In fact, the experimental evidence from a variety of studies, including surprisingly many that are cited favorably by M&S, suggests that people do not have these particular skills. One general challenge in extracting broader lessons from experimental data is that the skills of understanding, evaluating, and producing arguments are vaguely defined in the literature in general, and the target article is no exception. There is a crucial distinction between argument content and argument structure that is ignored, and some studies focus solely on argument content, while others focus on argument structure. The extent to which either kind of study supports claims about participants' ability to reason well depends on this distinction in an important way.

The definition of an argument given by M&S is standard: A set of statements, one of which is the conclusion, which is supposed to be epistemically supported by the other statements, called the *premises*. The *content* of an argument refers to the propositions that are expressed by the premises and conclusion, whereas the *structure* of the argument refers to the way the premises work together to support the conclusion. Successfully understanding an argument consists in being able to identify both the content and the structure of the argument: the conclusion, the premises, and the particular way the premises support the conclusion (e.g., whether the premises are linked or convergent). Successfully evaluating an argument consists in being able to assess the content (i.e., determine whether the premises are true) and the structure (i.e., determine whether, assuming that they are true, the premises actually do support the conclusion). Finally, successfully constructing an argument consists in being able to supply true premises and specify how those premises work together to support the conclusion. Although structure and content are both relevant for all three activities, they are relevant in different ways, and so great care is required (but not always taken) in designing experimental tasks that appropriately test them.

Problematic empirical evidence arises for all three: argument understanding, argument evaluation, and argument production. For the first process, there actually seems to be scant research in the area of argument understanding. The little research that does exist in this area is mixed. Some studies (e.g., Ricco 2003, cited by M&S) suggest that for simple arguments, adults can, when prompted, differentiate between linked and convergent arguments. Other studies, however, suggest that, even for simple arguments, untrained college students can identify the conclusion but without prompting are poor at both identifying the premises and how the premises support the conclusion (Harrell 2006; 2008; forthcoming).

Second, argument evaluation is usually loosely, and only implicitly, defined as being able either to identify reasoning fallacies or to differentiate reasonable arguments from unreasonable ones. The research on argument evaluation seems mixed, at best. In particular, a number of systematic biases have been found. When witnessing an argument from the outside, participants' judgment of the burden of proof depends on who speaks first (Bailenson & Rips 1996, cited by M&S), and participants routinely mistake innocuous repetition for circularity (Rips 2002, cited by M&S). When participating in an argument themselves, participants tend to reason less well than when witnessing an argument (Neuman et al. 2006; Thompson et al. 2005b; both cited by M&S).

Finally, in many of these studies, the perception by the researchers that participants were able to "build complex arguments" (sect. 2.2, para. 3) is vague or ambiguous. Producing an argument is importantly different from, for example, mere fact gathering, but the research focuses almost exclusively on nothing more complex than the listing of reasons to believe. Even for this simple kind of argument production, studies suggest that both low- and high-cognitive-ability participants have difficulty producing evidence for a claim (Sá et al. 2005, cited by M&S).

Contrary to the claims by M&S, a wide literature supports the contention that the particular skills of understanding, evaluating, and producing arguments are generally poor in the population of people who have not had specific training and that specific training is what improves these skills. Some studies, for example, show that students perform significantly better on reasoning tasks only when they have learned to identify premises and conclusions (Shaw 1996, cited by M&S) or have learned some standard argumentation norms (Weinstock et al. 2004, cited by M&S). M&S may be correct that some of these negative results arise because the stakes are too low, but many studies that show improvements from specific training occur in high-stakes environments like a college course (Harrell, forthcoming; Twardy 2004; van Gelder 2005; van Gelder et al. 2004). This suggests that difficulty with understanding, evaluating, and producing arguments may be a deeper feature of our cognition.

## The argumentative theory of reasoning applies to scientists and philosophers, too

doi:10.1017/S0140525X10002931

John A. Johnson

Department of Psychology, Pennsylvania State University, DuBois, PA 15801.

j5j@psu.edu

<http://www.personal.psu.edu/~j5j/>

**Abstract:** Logical consistency demands that Mercier and Sperber's (M&S's) argumentative theory of reasoning apply to their own reasoning in the target article. Although they hint that their argument applies to professional reasoners such as scientists and philosophers, they do not develop this idea. In this commentary, I discuss the applicability of argumentative theory to science and philosophy, emphasizing the perils of moral reasoning.

Mercier and Sperber (M&S) argue that the primary evolved function of reasoning is persuasive argumentation. If the primary function of any evolved trait – including reasoning ability – is the same for all members of a species, then it follows that professional reasoners (including scientists and philosophers) are primarily in the business of persuasive argumentation. Furthermore, if M&S's dual-process model of reasoning is accurate, professional reasoners initially arrive at their conclusions by intuitive leaps and only later construct logical arguments to convince others of these conclusions. The notion that scientists and philosophers are more concerned with persuading others that something is true than with discovering truth contradicts the image of scientists and philosophers as dispassionate truth-seekers. This response to M&S's target article aims to develop this subversive implication of their argument.

That M&S's argumentative theory applies to their own reasoning is necessary if their theory is to be consistent. To suggest otherwise is to commit what Little (1972) called the *nonreflexive fallacy*. Yet M&S spend virtually the entire article discussing studies of nonscientists and nonphilosophers, with just the briefest mention of how their theory might apply to professional reasoners. One exception is a reference to reviewers of scientific manuscripts who look for flaws in papers to justify rejection when they do not agree with a paper's conclusion. They also remark near the end of their article that even among scientists the ability to control one's own biases is "uncommon" and "almost freakish" (sect. 6, para. 7).

Perhaps the dearth of examples of professional-reasoning-qua-argumentation is due to space limitations. Or, perhaps there is little empirical research on this topic. Or, perhaps other professional reasoners will not find the theory as congenial as M&S suggest in their concluding sentence. After all, it could be somewhat demeaning to see one's professional activity (reasoning) as equivalent to ordinary squabbling over whether my favorite sports team is better than your favorite sports team. Whereas Little (1972) aims to elevate ordinary people to the status of scientists, M&S appear to be challenging the status of scientists and philosophers as elite thinkers. To suggest that "[s]killed arguers, however, are not after the truth but after arguments supporting their views" (see the M&S abstract) is to challenge the idea that scientists and philosophers are motivated in an unbiased way by pure curiosity about what is true.

I believe that we professional reasoners should accept M&S's humbling view of our activities because it is an accurate description of reality. Yes, we are interested in truth, but we relish the thought of convincing others that we have discovered important truths. I must confess that the M&S thesis was immediately congenial to me because it affirms my own long-held beliefs about how professional reasoners such as scientists and moral philosophers go about their work (Johnson et al. 1988). Observations of the actual behavior of scientific researchers indicate that textbook descriptions of science are highly inaccurate. Scientists do not begin with a thorough set of dispassionate observations about

some aspect of the world, followed by formulation of hypotheses that are tested and immediately abandoned if disconfirmed by data. I propose that the following account is more accurate.

Research for professional reasoners begins with an emotional attraction to certain ideas, an attraction Tomkins (1965) called “ideo-affective resonance.” This emotional resonance can cause scientists to cling tenaciously to ideas, even in the face of counter-evidence. In some ways, science resembles legal proceedings in which the very best case for guilt or innocence is presented by uncompromising prosecuting and defense attorneys, respectively. Scientists who resonate to different views clash in conferences and in journals. Each seeks to convince others that he or she is correct. M&S review research indicating that when members of groups holding disparate views debate, each arguing for a different view, “truth wins” (sect. 2.3, para. 1). Perhaps truth does win often enough in jury trials and scientific debates, but as we all know, sometimes it does not. M&S might be expressing unwarranted optimism here.

I want to close my commentary with some observations about moral reasoning. Research by Haidt (2001), mentioned by M&S, and by Joshua Greene (2003) strongly supports a dual-process model wherein people instantaneously decide if an act is “good” and therefore something we “ought” to do by taking note of the immediate, reflexive feelings that emerge when thinking about the act. In the second stage of the dual process, they may attempt to defend their feelings in terms of rational argument. Professional philosophers are much better at the reasoning part of the process, but are still guided initially by emotional reflexes. The immediacy and inevitability of certain emotions (e.g., revulsion on contemplating the torture of a child) can lead philosophers and nonphilosophers alike into making pronouncements such as “That we ought to refrain from torturing children is a moral truth.”

But only propositions about *what is the case* can be true or false. Moral pronouncements express reflexive feelings about how we ought to behave and are therefore not truth-apt. “Moral truth” is a category mistake. I have a yet-untested two-part hypothesis about why so many people (including moral philosophers) make this apparent category mistake (Johnson 2007). First, human beings are prone to mistakenly assuming that when they feel a strong and immediate emotion, this is a reliable sign of a self-evident truth. Second, although moral systems evolved because they conferred benefits on all participants (compare M&S’s observation that persuasive communication must be sufficiently beneficial to both parties, else the capacity for being persuaded would be selected against and go out of existence), the propensity of a person to be responsive to moral “oughts” can be exploited by someone who benefits at that person’s expense. Compare, for example, the persuasiveness of “Give me ten percent of your money because I want it” with “That we have a duty to tithe to the church is a venerable moral truth.” Scrutiny of any rhetorical effort is wise, particularly those in the moral domain.

## True to the power of one? Cognition, argument, and reasoning

doi:10.1017/S0140525X10002992

Drew Michael Khlentzos and Bruce Stevenson

Language and Cognition Research Centre, Psychology, School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale 2351, Australia.

dkhlentz@une.edu.au    bstevens@une.edu.au

<http://www.une.edu.au/staff/dkhlentz.php>

<http://www.une.edu.au/staff/bstevens.php>

**Abstract:** While impressed by much of what Mercier & Sperber (M&S) offer through their argumentative hypothesis, we question whether

the specific competencies entailed in each system are adequate. In particular, whether system 2 might not require independent reasoning capabilities. We explore the adequacy of the explanations offered for confirmation bias and the Wason selection task.

For Mercier and Sperber (M&S), what appears as poor reasoning is actually appropriate argument – social dialogue facilitates reasoning by prompting agents to formulate arguments and defend them from objections. M&S propose a dual-process model with system 1 ( $S_1$ ) a consortium of inference mechanisms and system 2 ( $S_2$ ), an  $S_1$  apologist. We identify some features we think require clarification and provide alternative interpretations of phenomena used by M&S to support their model.

If  $S_1$  generates conclusions without revealing their derivation (modular-like), then where does  $S_2$  acquire the competence to support these arguments? What type of reasoning is required for it to construct these arguments, or does it run data back through  $S_1$  for a reasoned result? Related to this is the issue of argumentative contexts which trigger  $S_2$ . These appear to be richer in information, creating a potential confound for the argumentative hypothesis: Is it the argumentative feature or the increased information that is critical?

The social psychology findings M&S adduce to support their view present a puzzle for it: How can truth win out amongst sophisticated  $S_2$ s committed *not* to discovering the facts but to defending  $S_1$ ’s representation of them? Convergence-on-truth suggests there’s more to  $S_2$  than defence of  $S_1$ . One alternative views  $S_2$  as a dynamic, defeasible reasoner that sifts through  $S_1$  outputs, *independently* generating conclusions to be updated in the light of new information.

Presumably  $S_1$  must support probabilistic as well as deductive inferences. In which case, *some* regulatory role for  $S_2$  is inescapable. Suppose  $S_1$  has both deductive and probabilistic mechanisms and these produce compatible results with input X both deductively entailing and probabilistically supporting Y. Imagine new evidence E emerging that undermines Y so that X + E makes Y *not* probable. Nonetheless, E cannot affect the derivation of Y from X. So X + E still entails Y. Whence  $S_2$  has to decide whether to defend Y since it is derivable from X + E or surrender Y as X + E makes Y improbable. How would it make this decision?

Consider now M&S’s views on *confirmation bias*. M&S deny confirmation bias is a flaw in reasoning. Yet if the aim of each agent’s  $S_2$  is to persuade others, confirmation bias would just polarize views with no agent prepared to listen to another’s arguments. Alternatively, if each  $S_2$  defends an agent’s beliefs against objections, amassing evidence for those beliefs is important but anticipating likely objections and preparing a defence is no less so. Relative to aims of persuasion or defence, then, confirmation bias registers as a fault in reasoning.

Compare an M&S-styled  $S_2$ -reasoner Aaron with a defeasible  $S_2$ -reasoner Belle. Aaron is convinced the river mussels are good to eat since he’s eaten them the past five days. Belle felt ill after eating them the day before. She advises Aaron to refrain. Aaron’s  $S_2$  considers positive evidence and discounts negative evidence. So Aaron eats the mussels and falls ill. In contrast, Belle’s  $S_2$  constructs fast generalizations on the fly. Having eaten them for four days, Belle inferred (G) *the mussels are good to eat*. But now her  $S_2$  enables Belle to adopt a position appropriate to the evolving evidence. The crucial difference between Aaron and Belle is this: Were they to swap roles, Belle would feel no internal pressure from *her*  $S_2$  to eat the mussels (unlike Aaron from his): Evidence someone else fell ill can prompt a defeasible reasoner to update (G) as disconfirming and confirming evidence are weighted equally. Whilst M&S’s model allows  $S_1$  to update information, *reasoning* to a new conclusion (belief revision) appears anomalous.

Does the *argumentative hypothesis* yield the best explanation of reasoning performance? Take the Wason selection task. M&S claim that when agents are asked to assess the truth of (W) *If there’s a vowel on one side of a card, there’s an even number on its other side* for an E, K, 4, 7 array, their  $S_1$  matches cards to

verbal cues, prompting them to choose the E card and 4 card. Their  $S_2$  then defends this choice against objections.

This matching hypothesis belies the semantic complexity of (W), which contains an indefinite “a card” nested within the scope of a conditional. Such constructions can vary in meaning. “If Ralph has a credit card, he buys tickets with it” is true if Ralph has four different credit cards but uses only one to buy tickets. However “if Sue has a pet, she feeds it” is false if Sue feeds her goldfish, starving her three kittens: “A pet” means *every* pet where “a credit card” means *some* credit card. Indefinites such as “a card” in (W) could be assigned a default existential reading (*some*) by an  $S_1$  semantic module. If contextual clues or background information don’t subsequently override it, this (mistaken) interpretation could become fixed, leading subjects to look for some card with an even number on one side and a vowel on the other, as the majority do. Alternatively, if the semantic module defaults to a *generic* reading of (W), since generics (like *the mussels are good*) tolerate exceptions, the 7 card’s role becomes opaque, rendering it non-salient once more.

For defeasible reasoners used to generic generalisations, figuring out the minimum number of occluded elements that need to be revealed to determine whether (W) is true is no trivial task. They fail not because they aren’t called on to defend their view, as M&S contend, but because they have little or no experience testing exceptionless generalisations.

Why then do they succeed in group settings? A rival problem-solving model of reasoning says this is due to two factors:

1. Group settings are informationally rich with alternative hypotheses articulated, reasons for them explained and discussed, *and*

2. Subjects’ semantic modules are specifically designed to compute the logical implications of *others’* statements to determine what those statements mean (Crain & Khlentzos 2010).

Further, since this model assumes subjects share a common goal of finding the correct solution, it is no mystery why subjects should converge on that solution.

## What people may do versus can do

doi:10.1017/S0140525X10002864

Deanna Kuhn

Department of Human Development, Teachers College, Columbia University, New York, NY 10027.

dk100@columbia.edu    www.educationforthinking.org

**Abstract:** It warrants examining how well people can come to argue under supportive conditions, not only what they do under ordinary conditions. Sustained engagement of young people in dialogic argumentation yields more than the temporary “contextual effect” that Mercier & Sperber (M&S) identify in the target article. If such engagement were to become the norm, who can say what the argumentative potential of future generations is?

In the target article, Mercier & Sperber (M&S) make the strong claim that epistemic goals are not well served by argumentative reasoning because it does not enhance the accuracy or validity of one’s claims. Evidence is too ample to deny that people commonly invoke argumentative reasoning to support their assertions in ways that are habitual, often mindless, and self-serving. Where I would fault M&S, rather, is in their focus on how argumentative reasoning most often does function, to the exclusion of how it can (and for some does) come to function, as a consequence of education, engagement, and a recognition of its epistemic value. Although people may use argument in self-serving ways that they are in limited command of, it doesn’t follow that they cannot achieve greater conscious command and come to draw on it in a way that will enhance their cognitive power.

Moreover, as my own most recent research clearly shows (Goldstein et al. 2009; Kuhn & Crowell, in press; Kuhn et al. 2008), sustained engagement of young people in dialogic argumentation yields more than the temporary “contextual effect” of a social setting that M&S identify. In their review of our research, they focus on earlier work in which on a single occasion participants are asked to generate an argument to support their view regarding the cause of a particular social problem (Kuhn 1991). The participants generally do poorly, with little sign of improvement from adolescence through old age, tending to describe a plausible scenario of how the problem could arise and failing to differentiate that scenario from actual or potential evidence that this is how it does in fact arise.

In more recent work, we have engaged young people in sustained dialogic argumentation about significant issues. We focus on middle school as an optimal period to undertake this effort, and we follow Billig (1987), Graff (2003), and, before him, the sociocultural tradition of Vygotsky (1978) and others, in taking the everyday social practice of argumentation as a starting point and pathway for development of individual argumentative skill. The dialogic context provides the “missing interlocutor” (Graff 2003) that gives expository argument its point. The medium of discourse is electronic, yielding the significant advantage of providing a transcript of the exchange that remains available throughout and following the discourse. Contributions to face-to-face discourse, in contrast, disappear as soon as they are spoken. In addition to serving as a reference point and framework during the dialogs, these transcripts become the object of various reflective activities participants engage in.

With sustained engagement over multiple months, dialogic argumentation progresses from what starts out as the norm among young adolescents – exposition of one’s own views with scant attention to those of the opponent – to recognition of the need to attend to the opponent’s claims and identify weaknesses, and from there to sustained sequences of counterargument and rebuttal. Perhaps most important, not immediately but with time, these newly developed skills transfer to the individual context that M&S focus on. Relative to a carefully matched comparison group at the same school, our participants wrote superior individual essays on a new topic, ones more often addressing strengths and weaknesses of both sides of an issue. Equally important is young people’s progress in the epistemological domain of recognizing the centrality of counterargument and of evidence to sound argumentation. Again relative to the comparison group, participants showed greater recognition of the role of both, for example in seeking information to bring to bear on their arguments.

In a word, we need to examine how argument may come to be used under these favorable, supportive conditions, not only under more ordinary conditions. If broader engagement of this sort were to become the norm, who can say what the argumentative potential of future generations is? With education systems worldwide claiming commitment to the broad goal of students learning to use their minds well, so as to be able to apply them to new, unknown problems, we should at least seriously explore the question.

## The world looks small when you only look through a telescope: The need for a broad and developmental study of reasoning

doi:10.1017/S0140525X10002918

Darcia Narvaez

Department of Psychology, University of Notre Dame, Notre Dame, IN 46556.

dnarvaez@nd.edu    http://www.nd.edu/~dnarvaez

**Abstract:** If the target article represents the summary findings of the field, reasoning research is deeply flawed. The vision is too narrow and

seems to fall into biological determinism. Humans use reasoning in effective ways apparently not studied by researchers, such as reasoning for action. Moreover, as the brain develops through adulthood and from experience so do reasoning capabilities.

My two critiques address the limited scope of the research and the neglect of human development. These undermine the generalizability of Mercier and Sperber's conclusions.

**First, the way reasoning is defined and studied leads to narrow, incomplete findings.** Mercier and Sperber cite research that ignores a great deal of reasoning behavior. For example, at the *sociopolitical* level humans use reason to design and change laws, constitutions, institutions, and visions such as the Declaration of Human Rights. Reasoning at the *everyday* level includes figuring out what course of action to take: for our ancestors, when to migrate to the next foraging ground; for us, how to balance the daily demands of work and family. Nor is there any reference to how people reason after a poor outcome: For our ancestors, why was the hunt unsuccessful today and what can we do differently tomorrow? For us, how did I lose my cool with my child and how can I avoid that in the future? The authors make no distinctions among types of goal-motivated reasoning, excluding pre-hoc (planning – what should my plans be today?), post-hoc (reflecting—how did things go?), and online executive reasoning (e.g., this plan is not working, what should I do?). Even children employ reasoning for action when they consider how to climb a tree, how it is going, and reflect on their failure or success.

The authors describe reasoning as a process more akin to rhetoric, completely leaving out practical reasoning. They claim that human reasoning evolved to competitively persuade others of one's viewpoint rather than for making the best decision. This astonished me – how adaptive would it be to follow a rhetorically gifted con man or inexperienced group member in the Pleistocene? The experience-based wisdom of the elders was much more advantageous.

The research tasks used and interpretations employed seem to presume that humans are primarily self-interested, a notoriously implausible view outside the West (Sahlins 2008). Of course there can be situations that press individuals to be competitive rather than cooperative in decision making, but from anthropological accounts our ancestors were cooperators within their groups, not the ego-driven competitors described by the authors (Fry 2006). It seems important to distinguish between self-interested cognition and cognition motivated by other concerns. For example, how do the authors explain the efforts of Warren Buffet and Bill Gates (givingpledge.org) to persuade wealthy individuals to contribute half of their wealth towards charity and the common good? Certainly they used rhetorical skills in their mission but whence the motivation? How would the authors explain the reasoning skills and motivations of the likes of Nelson Mandela and Abraham Lincoln in solving their society's challenges?

**Second, the authors seem to assume that people don't develop in reasoning capacities and that the college student represents human capability.** There seems to be an implicit biological determinism in the target article, a view that is empirically untenable (Lewontin et al. 1987).

The research findings are circumscribed by the population usually studied – college students – giving a false impression of human capabilities. Wisdom is found more typically in mature adults, not sophomores. Brain development after the teenage years is fundamental for mature reasoning capabilities. In the mid to late 20s humans acquire greater executive function capacities (Luna et al. 2001), which allow for the ability to move past the subcortical decision-making system, highly influenced by the situation, and use prefrontal capacities that facilitate perspective taking and empathy with greater awareness of consequences (Goldberg 2001). In middle age, adult brains undergo further myelination, peaking in inductive reasoning (Schaie & Willis 2010).

One cannot draw any firm conclusions about reasoning without examining mature adults in ecologically valid tasks. Researchers should study reasoning in adults as they perform their roles as experts: experienced parents, judges, ministers and counselors, shopkeepers and community leaders, umpires and zookeepers. These experts learn to merge self and moral interests or they falter in their roles. Experts develop in reasoning capabilities, tapping into intuitions, explicit knowledge, and domain-specific paradigms that novices lack (Hogarth 2001). Instead, the focus in psychological studies seems to be on what underdeveloped minds and brains of a certain sort do well – make quick judgments and use words to manipulate others to get one's way. Elsewhere I criticize this shortsighted focus in moral judgment research (Narvaez 2010).

Further, it's not at all clear that the researchers are studying optimal brains even at the college level. The prefrontal cortex, the seat of executive functions, apparently can be damaged prior to its maturation from addictive use of drugs (Bechara 2005) and activities that keep the more primitive parts of the brain active, such as violent video games (Mathews et al. 2005), suggesting that reasoning capacities may be diminished in those who engage deeply in such activities. Sociocultural factors also affect reasoning, such as deteriorating child-rearing practices (Narvaez 2008), which may play a role in the lower rates of empathy (Konrath et al., in press) and moral reasoning (Thoma & Bebeau 2008), and in greater egocentrism if not the narcissism (Twenge & Campbell 2009) reported in college students today.

Finally, it is highly questionable whether it is appropriate at all to generalize to human nature from the study of westerners or Americans. Henrich et al. (2010) point out how the vast majority of psychological studies and conclusions are based on Western, Educated, Industrialized, Rich, and Democratic (WEIRD) participants who represent less than 12% of the world population (college students, a subset of that).

The review leaves this reader unsatisfied with the work in the field. Reasoning needs to be defined more systematically and holistically by those who study it. In light of the narrow definition, the limited task set, and the population usually studied, it is not surprising that the findings are so pessimistic. Humans use reason in many more adaptive ways than described here. People and brains develop; experience and culture matter. Rather than a review of human capabilities, we have a glimpse into a narrow slice of reasoning by immature reasoners from an abnormal culture.

## Putting reasoning and judgement in their proper argumentative place

doi:10.1017/S0140525X1000289X

Mike Oaksford

Department of Psychological Sciences, Birkbeck College, University of London, London WC1E 7HX, United Kingdom.

mike.oaksford@bbk.ac.uk

<http://www.bbk.ac.uk/psyc/staff/academic/moaksford>

**Abstract:** This comment agrees with Mercier and Sperber's (M&S's) thesis on the argumentative function of reasoning but suggests that an account of argument strength is required. A Bayesian account of argument strength (Hahn & Oaksford 2007) shows how the deployment of deductive fallacies, weak inductive arguments, and judgment fallacies such as base-rate neglect, can all be rationally defended in the right argumentative context.

Mercier and Sperber's (M&S's) hypothesis – “the function of reasoning is argumentative. It is to devise and evaluate arguments intended to persuade” (see their abstract) – is a timely and important thesis that sits well with related arguments

in the literature. For example, Hahn & Oaksford (2007) argued:

Typically, reasoning takes place in the service of argumentation, that is, in the attempt to persuade yourself or others of a particular position. Argumentation is the overarching human activity that studies of deductive reasoning, inductive reasoning, judgment, and decision making are really required to explain. (p. 705)

M&S argue that demonstrations “that people are skilled at assessing arguments seems to stand in sharp contrast with findings from the psychology of reasoning” (sect. 2.1, para. 3). They discuss participants’ failures to employ logical rules but not the frequent endorsement of logical fallacies, although they suggest that, “unlike logical fallacies, fallacies of argumentation come in degrees” (sect. 2.1, para. 4). Many argumentation theorists argue that all argumentation, not just the fallacies, is a matter of degree. For example, Perelman and Olbrechts-Tyteca (1969) argued, “The aim of argumentation is not to deduce consequences from given premises; it is rather to elicit or increase the adherence of the members of an audience to theses that are presented for their consent” (p. 9). Here we argue that logical fallacies and logical rules subservise argumentation but only to the extent that they to “come in degrees.” A corollary to this argument is that judgmental fallacies such as base-rate neglect may also arise in the same context.

I focus on the deductive fallacy of *denying the antecedent* (DA) in conditional reasoning, for example:

If a bird is a swan, then it is white  
That bird was not a swan  
Therefore That bird was not white

This instance of DA is clearly a logical fallacy as there are white birds that are not swans. However, suppose someone (*Con*) is deploying DA against someone (*Pro*) who has argued that a particular bird was white via the logical rule of modus ponens (MP), *if a bird is a swan, then it is white, that bird was a swan, therefore, it was white*. To refute *Pro*’s argument the respondent must deny one of the premises. *Con* chooses to “deny the antecedent” – that is, to deny that the bird was a swan, from which it “follows” that the bird was not white (Godden & Walton 2004). From an argumentative perspective this seems like a perfectly sound strategy.

However, the falsity of the consequent – the bird was not white – could not follow logically from this use of DA. Rather it undermines another property of *Pro*’s intended conclusion – that is, “that it should be believed” (Godden & Walton 2004, p. 232). A Bayesian probabilistic approach to argument strength for the conditional cashes out this intuition (Hahn & Oaksford 2007; Oaksford & Hahn 2007). Hahn and Oaksford (2007) proposed that people’s degree of belief in the conclusion given the premises provides one measure of argument strength. They equate this measure with the conditional probability of the conclusion given the premises: for MP this is  $\Pr(\text{bird is white}|\text{bird is a swan})$  (see Oaksford & Chater 2007; Oaksford et al. 2000). Another measure, which they refer to as “argument force,” is the change in degree of belief brought about by the argument. They equate this measure with the likelihood ratio, which maps the prior odds into the posterior odds: for MP this ratio is  $\Pr(\text{bird is white}|\text{bird is a swan})/\Pr(\text{bird is white}|\text{bird is not a swan})$ .

It is quite easy to show that contexts can arise where both *Pro*’s and *Con*’s arguments seem warranted. Suppose that the swans are in a bird sanctuary containing equal numbers of white and black swans:  $\Pr(\text{bird is white}|\text{bird is a swan}) = 0.5$ , and that most of the birds in the sanctuary are neither white nor swans: for example,  $\Pr(\text{bird is white}) = \Pr(\text{bird is a swan}) = 0.1$ . On this distribution, the likelihood ratio for the MP argument is 9; the bird is 9 times more likely to be white given it is a swan than if it were not a swan. In contrast, the likelihood ratio for

DA,  $\Pr(\text{bird is not white}|\text{bird is not a swan})/\Pr(\text{bird is not white}|\text{bird is a swan})$ , is much lower; the bird is only about 2 times more likely not to be white given it is not a swan than if it were a swan. So, in terms of the *force* of the respective arguments, MP seems stronger than DA, which seems to warrant *Pro*’s opening argument.

*Con*’s DA counter-argument can be characterised as noting that priors matter and that it is highly unlikely that the bird was a swan. That is, his counter-argument suggests that *Pro* has neglected the base rates, a common bias in judgement research (Bar-Hillel 1980; Kahneman & Tversky 1973). When priors are taken in to account, *Pro*’s MP argument is weak; that is, the posterior probability is only 0.5. In contrast, the posterior probability of the DA argument,  $\Pr(\text{bird is not white}|\text{bird is not a swan})$ , is 0.94. So, in this context, in terms of the *strength* of the respective arguments DA seems stronger than MP, which seems to warrant *Con*’s counter-argument: The conclusion of the DA argument should be believed more than the conclusion of the MP argument. The next exchange between *Pro* and *Con* would probably focus on *Pro*’s grounds for apparently neglecting the base rate and believing that for this particular bird,  $\Pr(\text{bird is a swan}) > 0.1$ .

This analysis supports M&S’s contention that reasoning is for argument but only when we give a probabilistic analysis of how logical rule and fallacies affect people’s degrees of belief in their conclusions. In conclusion, we would argue that putting reasoning and judgment in their proper argumentative place compels us to a probabilistic understanding of deductive reasoning.

## On the design and function of rational arguments

doi:10.1017/S0140525X10002943

John E. Opfer and Vladimir Sloutsky

Department of Psychology, Ohio State University, Columbus, OH 43210.

opfer.7@osu.edu sloutsky@psy.ohio-state.edu

<http://faculty.psy.ohio-state.edu/opfer/opfer/opfer.html>

<http://cogdev.cog.ohio-state.edu>

**Abstract:** It is unclear how an argumentative environment would select for better reasoning given three general findings. First, argument rationality typically fails to persuade poor reasoners. Second, reasoned argumentation competes with more persuasive and less rational arguments for limited cognitive resources. Third, those poor at reasoning fail to distinguish between valid and invalid arguments. Reasoning, therefore, is poorly designed for argument.

Did reasoning evolve because it is best suited for communicating arguments? In the target article, Mercier & Sperber (M&S) claim this might be the case even though individuals poorly generate and recognize valid arguments. If, however, individuals are poor at generating and recognizing valid arguments, there are several – in our view, insurmountable – obstacles for this type of thinking evolving for any social function.

First, *reasoned argumentation would fail to achieve the goal of persuasion – changes in beliefs and attitudes*. One of the most common pitfalls of everyday reasoning is that nonrational factors (e.g., believability of an argument, confidence and status of the arguer, and self-interest of the receiver) trump logical validity in argument evaluation (for review of believability effects, see Evans et al. 1993). The emergence of a trait for recognizing valid arguments, therefore, would offer no particular benefit in a context of like-headed reasoners: Valid arguments would only convert the converted. Examples of this come from the phenomenon of *belief polarization* (Lord et al. 1979), where open exchanges of arguments serve only to pull disputants away from recognizing valid alternatives.

Another example of valid reasoning failing to persuade comes from studies of dyadic learning in children. A common finding in this literature (e.g., Ames & Murray 1982) is that when conservers and nonconservers share their reasons for conservation problems, the former persuade the latter. But is this a case of reason triumphant – or of less-confident students yielding to their more-confident peers? Suggesting this might be a victory for confidence over reason, Levin and Druyan (1993) found that confidence was much higher among conservers than nonconservers. Further, when problems were switched to science problems with high rates of misconceptions, the competence/confidence correlation was flipped – and children found the confident incompetents more persuasive than their better-reasoning peers. (From this perspective, it's easy to see why dogmatic creationists want scientists to “teach the controversy.”)

Could it be (as argued by M&S) that reasoning evolved to help people detect untrustworthy sources by flagging inconsistency in their arguments? Developmental evidence suggests that this is unlikely because children detect trustworthiness long before they detect argument inconsistency. For example, when Morris and Hasson (2010) presented children with simulated arguments of the simplest logical form (e.g., Puppet 1 claims “There is a sticker in the box!” and Puppet 2 claims “There is *no* sticker in the box!”), nearly 100% of 4- and 5-year-olds failed to detect an inconsistency between the two claims (Morris & Hasson 2010). At the same time, ability to detect a source as trustworthy emerges by 3 or 4 years of age (e.g., Jaswal & Neely 2006; Koenig et al. 2004). Given this, it cannot be that detecting trustworthiness requires the ability to detect argument inconsistency.

Second, *reasoned argumentation is expensive and may compete for limited cognitive resources with less expensive “hot” cognition*. Recognizing and generating valid arguments (that are not already believable) requires substantial cognitive resources, including a heavy cost to working memory (Barrouillet et al. 2000). This cost increases dramatically with an increase in the number of premises and introduction of quantifiers. At the same time, if the goal is persuasion, “hot” cognition (e.g., appeals to emotionally laden examples) can offer a less expensive alternative (Petty & Cacioppo 1996). Given that the cost of generating and comprehending logical arguments outweighs those of “hot” cognition, without offering substantial benefits, a mutant with an elevated ability for logical argument would have no competitive advantage against her demagogic disputants. Thus, it is difficult to see how the argumentative context would provide the ideal environment for the evolution of logical argument.

Third, *operator/receiver parity precludes benefits of social learning or knowledge transfer*. Although it is often tempting to draw analogies between language and reasoning (e.g., Braine & O'Brien 1998), the difference between the two is profound. Unlike reasoning, language proficiency is (more or less) universal in human adults. Consequently, in linguistic communication, more proficient language users (operators) can pull less proficient language learners (receivers) along the path to greater proficiency. This is not the case with reasoned argumentation, however. Here, operator and receiver characteristics are more symmetrical: When a person who is receptive to invalid arguments is put in the role of the operator, invalid arguments follow, and when a person who produces invalid arguments is put in the role of the receiver, valid and invalid arguments are not discriminated. Consequently, communicating reasons across individuals cannot, by itself, add anything to argumentation. Indeed, one of the most striking findings in cognitive development concerns how greatly change in language proficiency outpaces that of logical thinking, as well as how little reasoned argumentation is affected by observing a more proficient reasoner (as observed by Levin & Druyan 1993).

The failures of everyday reasoning that we think would make the argumentative context an inhospitable environment for the

evolution of reasoning are seen in a strange light by M&S. According to them, these failures support their account because they arise mostly *outside* an argumentative context. Yet, even if we were to stipulate this as true, superior reasoning in an argumentative context does not support their claim about the evolution of reasoning: It would imply that arguments facilitate reasoning, not that reasoning facilitates arguments. Yet, if reasoning is designed for arguments, as M&S contend, quality of reasoning *must* facilitate or hinder quality of arguments, whereas the reverse is unnecessary. To take M&S's analogy, to show that the structure of the foot is designed for walking, one must show that a different structure would facilitate or impede walking – not that process of walking strengthens the foot of the walker.

In our view, the kind of argument that is optimally designed for social communication – that is, would have the largest effect on manipulating another's behavior – is not necessarily a reasoned argument but an argument that is believable, emotive, easy to generate on the fly, and clear to others. Put simply, reasoned argumentation is no more likely to have evolved for social communication than is the posture of the foot to have evolved for disco.

## What is argument for? An adaptationist approach to argument and debate

doi:10.1017/S0140525X1000302X

David Pietraszewski

Department of Psychology, Yale University, New Haven, CT 06520-8205.

david.pietraszewski@yale.edu

**Abstract:** A consideration of selection pressures on the psychology of argument suggests that fixing the truth value of claims is not the primary criterion for argument generation or evaluation. Instead, argument psychology is designed to change representations in other minds as a way to negotiate conflicts of interest and as a way to signal social coordination.

Mercier and Sperber's (M&S's) analysis of reasoning as designed for argumentation represents another blow to certain long-held assumptions about cognitive processes: That reasoning is the abstract application of a propositional calculus used to determine what is true. Instead, M&S suggest that reasoning is the output of argumentation psychology, a suite of cognitive systems designed to handle incommensurate representations between people. This is courageous and provocative because it suggests that enterprises such as science are handled by a psychology designed for argumentation. Insofar as *reasoning* can be defined as an interestingly coherent natural category, M&S are likely correct.

However, the argument can be taken further. If reasoning is for argument, what is argument for? While M&S allude to this, there is some value in explicitly addressing the function of argument because it directly speaks to how argumentation psychology should work.

Consider the case of the evaluation of factual or policy claims. It is tempting to think that argument's proper domain in such cases is to determine the truth or accuracy of incommensurate representations – a natural consequence of information being distributed nonuniformly across bodies – that some people have access to information that others do not, and that, given imperfect information, each person is expected to prefer one's own data and conclusions and be wary of others. On this view, even if reasoning is for argument, then the ultimate logic of argument is the same as the classical view of reasoning – as a way of determining truth – albeit in a way that diverges from a rational view, by virtue of the division of information access in the real world.

However, a consideration of the selection pressures for argument suggests that this is not a complete picture of the function of argument and thus not a complete picture of the psychology of argumentation, even when dealing with claims ostensibly about truth. Here's why.

Communication exists because it affects behavior. This makes communicated information a particularly powerful way to leverage others' capacities. This creates opportunity for collaboration, as well as exploitation, and as such there will be selection on both senders and receivers of argument. M&S do not ignore that conflict of interest is inherent in communication – suggesting the existence of epistemic vigilance for ensuring that only “reliable,” “effective,” “quality” information is acted on. But what constitutes reliable, effective, or quality information? Certainly, as M&S imply, a criterion of accuracy applies to a class of claims. “The bear is on *this side* of the river” and “The bear is on the *other side* of the river” are factual claims warranting investigation of truth value. The bear is somewhere, after all, and argument will help determine where.

However, while there is reason to think that there is strong selection for discovering the accuracy for certain classes of claims, there is also good reason to think that there is little selection for searching for the accuracy of many others. Instead, if signaled information is capable of causing others to act, there is selection to broadcast representations that will cause others to modify their current or future actions with respect to the sender. Because utility and accuracy are ultimately separable, even for factual claims, there is not always selection for accuracy, but instead for a psychology that fixes representations – in oneself and in others – along actuarially beneficial dimensions.

This suggests at least two broad classes of selection pressures and, subsequently, two broad classes of argument psychology:

First, is a class of argumentation psychology designed to handle conflicts of interest over self-regarding and other-regarding behaviors, the goal of which is to change representations of the payoff structure of pursuing certain future plans. This view of argument has already uncovered a highly-specific “grammar” of argument structure in both senders and receivers (Petersen et al. 2010; Sell 2006; Tooby et al. 2008).

Second, is a class of argumentation psychology designed around social coordination. Because shared mental content is a consequence of coordinated activities, and because cooperation requires a meeting of the “minds,” shared mental representations can be used as markers and facilitators of social coordination. In other words, the exposition of claims – and the signal of agreement or disagreement with them – can be used as a social instrument to mark affiliation. Agreement and disagreement therefore become commodities in themselves as a way of signaling the coordination strength and challenging others. This class of argumentation psychology should be designed to conflate evaluations of the argument with the source and social context of the argument; who is arguing should be just as important as what they are saying when considering the “goodness” of an argument. Additionally, the motivation to argue, and the choice of argument content itself, should be at least in part the result of strategic non-conscious assessments of the local social world. This view of argument has already led to the discovery of evidence that the mind treats certain classes of claims as markers of social affiliation (Pietraszewski et al., in preparation).

These are not aberrant uses of argument, but, rather, these functions lie at the core of how the human psychological competence of argument – and thus how “reasoning” – works. The valuation of social coordination, for example, is likely built right into the sinew and bone of argumentation – both in terms of the criteria for generating arguments and for the criteria of assessing argument. This suggests that reasoning is not simply based on argument, but on argument designed for negotiating conflicts of interest and signaling social coordination rather than exclusively fixing truth.

## You can't have your hypothesis and test it: The importance of utilities in theories of reasoning

doi:10.1017/S0140525X10002980

Fenna H. Poletiek

Department of Psychology, Leiden University, The Netherlands.

poletiek@fsw.leidenuniv.nl

**Abstract:** Mercier and Sperber's (M&S's) theory of reasoning cannot predict reasoning in the absence of an argumentative context. Applying the theory to hypothesis testing behavior, I propose that hypothesis testing is often motivated by determining the true inference and that reasoning models should account for utilities (affected by various motives, including the wish to convince) of reasoning outcomes.

Mercier and Sperber's (M&S's) argumentative theory of reasoning (ATR) claims that reasoning is aimed not at improving knowledge, but at persuading. According to ATR, an argumentative context (actual or proactive) is a necessary condition for reasoning and reasoning is biased toward winning the argument. It will be argued that the very logic of ATR is problematic and that it can not deal with a large majority of reasoning contexts in which agents (also) reason to determine a true or valid conclusion. I propose that a theory of reasoning should incorporate *utilities* of reasoning outcomes, to explain various motives for reasoning, including argumentative motives.

Although M&S discuss in detail the *function* of reasoning, it is not always clear how reasoning is defined. If we assume reasoning to be about opinions and preferences for which no normative standard is available, an argumentative view is quite straightforward. People indeed may argue about political opinions and preferences with no other goal than to convince. However, if reasoning is the treatment of information about some (yet unknown) true or valid inference, with the objective to reduce inference errors, then the argumentative theory is puzzling. Indeed, the ATR theory of reasoning *disregards* this inherent concern of approximating the valid conclusion, going against what reasoning is defined to be.

The uneasy coexistence of finding out the truth and seeking confirmation for one's belief in ATR is apparent in M&S's analysis of the *confirmation bias* in hypothesis testing studies (Wason 1960). On the one hand, M&S acknowledge, in line with classical critical analyses of the confirmation bias (Klayman & Ha 1987; Poletiek 1996; 2001; Wetherick 1962), that participants' behavior in these studies is *not* indicative of a biased search of supporting test outcomes, but that it reflects a *sound heuristic*. This heuristic is testing hypotheses with *positive* predictions. Using M&S's example, suppose I believe that my keys are in my jacket because that is where I remember putting them. I look for them in my jacket (positive testing) and not in my purse (negative testing). Hence, as opposed to the interpretation of *positive testing* as a tendency to confirm (confirmation bias) (Cowley & Byrne 2005; Wason 1960), positive testing may occur with no other goal than finding out the truth, as the real-life example suggests. According to ATR, positive testing is a default heuristic that involves *no reasoning proper*.

However, as M&S further argue, reasoning can be triggered in hypothesis testing situations if an argumentative context is provided. Moreover, in such a context, reasoning is directed at falsification, according to ATR: Falsification is accessible provided that the situation encourages participants to argue against a hypothesis that is not their own, as M&S propose. This logic reveals the old misinterpretation that test *choice* (positive or negative) is tuned at the emergence of favorable test *result*. In fact, putting one's idea to a (either positive or negative) test *assumes* the objective to find out the truth and is therefore at odds with testing in order to save our idea from confirmations or disconfirmations. Poletiek (1996) showed that participants in a hypothesis testing experiment are aware of the incompatibility of finding out the truth and coming up with confirming test outcomes only. Participants felt that they were looking for information about the validity of a hypothesis, and that they could not control the test outcomes by choosing a test strategy. It was only



Table 1 (Poletiek). *Utilities of reasoning outcomes about H (the reasoner's subjective belief), assuming an epistemic motive (A), an argumentative motive (B), and both types of motives (C).*

A		
	H is true	Not-H is true
Infer H	10	0
Infer not-H	0	10
B		
Infer H	10	10
Infer not-H	0	0
C		
Infer H	10	4
Infer not-H	0	6

their *expectation* of the outcome that changed when they tested a hypothesis they did not believe to be true (falsification was expected). When testing their best guess, they expected a confirming outcome.

Besides determining the truth, many considerations may affect how we reason and whether we reason. These considerations can be modeled as a cost-benefit analysis of making the right inference. In Wason's (1960) task, students participating in the experiment might not care much about making reasoning errors. The key seeker might perform a negative test (looking for the keys in some place other than the jacket) because the costs of a positive one are too high: The jacket was left behind in a friend's house and the effort too great to walk back to check the jacket. Alternatively, we might be in a hurry and have time to check just one location, enhancing the benefits of good reasoning.

To predict reasoning behavior, we need a model with utilities of reasoning outcomes. A suitable tool is signal-detection theory. Motives can be expressed in the utilities of inference errors and in reasoning behavior predicted on the basis of the risk we are prepared to take with regard to particular erroneous inferences (Poletiek & Berndsén 2000). For example, as shown in Table 1, a pure epistemic motive would be modeled with (A) low utilities of making any false inferences. A pure argumentative motive would be expressed in (B) a very high willingness to make the false inference that our favorite hypothesis is true; and (C) reasoning with both motives (searching for a valid inference within both practical or argumentative constraints) is reflected with some in-between utilities with regard to making a false or a correct inference about our favorite hypothesis.

In this manner, reasoning in a variety of contexts and with a variety of goals can be modeled, offering an elegant alternative to the paradox of the ATR that we start up a reasoning trajectory about a prior belief *if and only if* the end of the route leads us to inferring that belief again.

### When reasoning is persuasive but wrong

doi:10.1017/S0140525X10002761

Robert J. Sternberg

Provost and Senior Vice President, Oklahoma State University, Stillwater, OK 74078.

Robert.sternberg@okstate.edu

**Abstract:** Mercier and Sperber (M&S) are correct that reasoning and argumentation are closely related. But they are wrong in arguing that this relationship is one of evolutionary adaptation. In fact, persuasive reasoning that is not veridical can be fatal to the individual and to the propagation of his or her genes, as well as to the human species as a whole.

In the target article, Mercier & Sperber (M&S) brilliantly show a crucial relationship between argumentation and reasoning. But the relationship is not the one they claim to show.

Consider two individuals facing an adaptive challenge. A threat of some kind is about to confront them. One of the individuals, A, recognizes the threat; the other, B, fails, for whatever reason, to recognize it. The two individuals argue over the existence of the threat or, perhaps, its severity. Each argues compellingly for his or her point of view. After all, with inductive reasoning based on incomplete and often ambiguous information, arguments can have inductive strength but not deductive certainty; and their inductive strength does not necessarily correspond to their veridicality (as anyone will have observed who has seen a defense attorney gets off his guilty client scot free). A and B both act on the basis of their reasoning. A survives and B dies (as a result of a bear attack, a lightning strike, an automobile accident, a plane crash, or whatever the threat in question happened to be).

A and B both used their reasoning in the service of argumentation, but reasoning was adaptive by virtue of the veridicality of its conclusion, not by virtue of the persuasiveness of the arguments (which may or may not correspond to veridicality in real-world ambiguous situations with incomplete information). So reasoning could scarcely have evolved in the service of argumentation, because those wonderful arguers who did not perceive things veridically would have been less able to reproduce than those arguers who did perceive things veridically. The brilliant reasoners who argued wrongly regarding threats had many more opportunities to perish before reproducing than those reasoners, persuasive or not, who saw threats as they were.

The same phenomenon occurs at a group level. Consider global warming. Global warming threatens the existence of human and other life on the planet Earth, and yet deniers, including scientists, put the life of all human organisms on the planet – the replication of the species' genes and hence the survival of the species – at risk. Reasoning is being used in the service of argumentation, but not always for evolutionarily adaptive purposes, at least with respect to the genes of the individuals involved.

The opening view of M&S that "reasoning should produce its best results when used in argumentative contexts, most notably in group discussions" (sect. 1.2, para. 11) is clearly wrong, as are the arguments that follow from it. The problem is that in this quotation, as throughout the article, there is an ambiguity regarding the meaning of "best results." If reasoning is about persuasiveness, perhaps the authors are right. But if reasoning is about veridicality, they are wrong.

Janis (1972) recognized the fallacy of the reasoning in his work on groupthink. He observed that groups of academically brilliant government officials could make horrible mistakes that were actually compounded by their being in groups. More generally, the phenomenon is referred to as "group polarization" (Moscovici & Zavalloni 1969). People like Robert McNamara and, more recently, Donald Rumsfeld, come to mind – people who, despite their academic brilliance, reasoned poorly, yet were able to persuade many by their (false) arguments. Stanovich (1993; 2009) coined the irrational but often persuasive reasoning of IQ-smart people "dysrationalia" (for related ideas, see also Sternberg 2002).

In the short run, reasoning in the service of argument may well be adaptive. For example, a job candidate who is persuasive is perhaps more likely to get the job than one who is unpersuasive; a politician who is persuasive is more likely to be elected or, at least, to be believed. But as recent presidential and other elections have shown, persuasiveness (at least to the masses) can be attained even by candidates who cannot string together a

coherent sentence. In the long run, the arguments of demagogic politicians are likely to lead to chaos. In the same way, the arguments of unscrupulous but persuasive clergymen are leading to terrorist attacks today.

If, as the authors argue, “the main function of reasoning is to produce arguments to convince others rather than to find the best decision” (sect. 1.2, para. 11), then human survival is in serious jeopardy. In today’s world, reasoning very likely is being used primarily to convince others rather than to find the best decision, but this use of reasoning is not evolutionarily adaptive for survival in the long run.

Perhaps, as a society, we are placing too much emphasis on reasoning in the service of argumentation, whether it is on college applications, in job interviews, or in elections. Instead, our society should place more emphasis on wisdom, the direction of reasoning (and other skills) toward a common good, over the long as well as the short term, through the infusion of positive ethical values (Sternberg et al. 2007). In a microcosm, the target article represents what has gone wrong with society as a whole: Society has come to care more about reasoning in the service of persuasion than reasoning in the service of truth or even some kind of ethical good. This trend risks leading not to better adaptation of humans but, rather, to their ultimate destruction.

## The chronometrics of confirmation bias: Evidence for the inhibition of intuitive judgements

doi:10.1017/S0140525X10002876

Edward J. N. Stupple<sup>a</sup> and Linden J. Ball<sup>b</sup>

<sup>a</sup>Centre for Psychological Research, University of Derby, Derby DE22 1GB, United Kingdom; <sup>b</sup>Department of Psychology, Lancaster University, Lancaster LA1 4YF, United Kingdom.

e.j.n.stupple@derby.ac.uk    l.ball@lancaster.ac.uk  
[http://psychology.derby.ac.uk/staff/Ed\\_Stupple.html](http://psychology.derby.ac.uk/staff/Ed_Stupple.html)  
<http://www.psych.lancs.ac.uk/people/LindenBall.html>

**Abstract:** Mercier & Sperber (M&S) claim that the phenomenon of belief bias – which they consider to be an archetypal manifestation of a general confirmation bias in human reasoning – provides fundamental support for their argumentative theory and its basis in intuitive judgement. We propose that chronometric evidence necessitates a more nuanced account of belief bias that is not readily captured by argumentative theory.

Mercier & Sperber’s (M&S’s) impressive argumentative theory reassesses the function of reasoning, not as involving the noble pursuit of truth, but instead as a Machiavellian communicative art with persuasion and self-interest at its core. A case in point is the infamous *confirmation bias*, whereby individuals seem motivated to seek confirmatory evidence for their existing beliefs and hypotheses and fail to look for counterevidence or counterarguments (Nickerson 1998). M&S claim that a quintessential demonstration of confirmation bias that supports their theory can be seen in the phenomenon of *belief bias*, where some contemporary theories suggest that people try to confirm believable conclusions but disconfirm unbelievable ones (see the selective processing model described by Evans 2007, and the multinomial model of Klauer et al. 2000). Thus, in both the case of believable and unbelievable conclusions, M&S claim that people show a motivation “to confirm their initial intuition” (sect. 3.3, para. 3) with unbelievable conclusions effectively promoting an intuition-guided *debias* (Evans 2000) leading to improved logical responding. M&S further propose that, when people deal with an unbelievable conclusion, “it is not that they reason *more* in this case. . . . It is just that the direction reasoning

takes is mostly determined by the participants’ initial intuitions” (sect 3.3, para. 3).

Our contention, however, is that this latter claim flies in the face of current chronometric evidence in the belief-bias literature, which suggests that a more subtle interpretation of the phenomenon is needed that is not couched purely in terms of the operation of a general confirmation bias. In particular, processing-time data for syllogistic arguments consistently indicate that participants reason *most* when the conclusion is *believable* but *invalid* (Ball et al. 2006; Stupple & Ball 2008; Thompson et al. 2003). Such a finding is inconsistent with M&S’s view that people simply seek support for prior beliefs. Confirmatory mental models of the premises of arguments with believable-invalid conclusions are readily available, so why should significantly longer processing times arise with these problems? We propose instead that many participants show a capacity to inhibit confirmation-oriented processing with such arguments, with the resulting attempt at logical analysis taking time to apply. Of course, the complexity of the required logical processing means that a belief-based response may still often win out, perhaps with people defaulting to such a response under cognitive load (cf. Quayle & Ball 2000). This would produce a response pattern that looks like confirmatory behaviour, but where the chronometric data support a very different account of the processing that is actually taking place.

To elaborate on our proposals we outline three possible routes that participants could take through Evans’ (2007) selective processing model when confronted with belief-oriented syllogistic arguments (cf. Ball 2010; Evans 2009). First, participants could employ a pure response bias and respond in accordance with belief without engaging any analytic processing whatsoever, either for a truth-seeking or argumentative purpose. Second, in accordance with argumentative theory, participants could seek confirmatory evidence so as to warrant the acceptance of believable conclusions (including believable-invalid ones) and the refutation of unbelievable conclusions (including unbelievable-valid ones). Finally, participants could attempt to suspend notions of belief and disbelief altogether. For example, rather than searching for a supporting model for an believable-invalid conclusion, they would inhibit a heuristic response as well as confirmatory-oriented analytic response, instead engaging in an exhaustive search for a model of the premises that provides a counterexample to the given conclusion.

The important question that follows from our analysis is this: What if examples of all three reasoning approaches were present in belief-bias data sets? We suggest that the resulting aggregation of these strategies would lead to the pattern of response times that has typically been observed (Ball et al. 2006; Stupple & Ball 2008; Thompson et al. 2003), with believable-invalid conclusions being associated with extended processing times relative to other problems because of the presence of a subset of reasoners who resist intuitive judgements. This group of reasoners would most likely be those described by Stanovich and West (2000) as adopting the “normative construal” of the task because they possess the cognitive capacity needed to reason through demanding deductive problems.

This latter interpretation of chronometric findings is supported by data that we have recently acquired (for a preliminary report, see Ball 2010) demonstrating that increased response times for believable-invalid problems are predictive of increased overall response accuracies across belief-oriented problems (i.e., these times reflect the behaviour of high-ability reasoners). These data also concur with the observation by Thompson et al. (2010) that the participants who spend more time reasoning are more likely to reach a logical conclusion. However, Thompson et al. propose an alternative interpretation of the inflated inspection times for believable-invalid problems that is more in keeping with argumentative theory than our account. In their Modified Verbal Reasoning theory they suggest that participants are motivated to find confirmatory support for believable

problems and so allow an extended response deadline for such problems (perhaps an attractive proposition for the argumentative theory). Thompson et al. claim that since the reasoning process is more difficult for invalid than valid problems, this combined with the increased motivation to support believable conclusions results in the corresponding increase in response times for believable-invalid problems. We would argue, however, that in adopting this perspective one would expect a far greater acceptance rate for believable-invalid problems than is observed (acceptances are typically in the 50%–70% range), since a confirming model is readily available to those participants who expend the effort trying to find one.

In sum, we disagree with M&S's assertion that a motivated attempt at confirmation necessarily drives belief-bias effects in syllogistic reasoning. Instead, we claim that many participants attempt to seek out the deductive truth of presented arguments and do so at the considerable cognitive expense of inhibiting their *intuitive judgements*, as is particularly evident in the case of syllogisms with believable-invalid conclusions. On this issue, at least, we would argue against M&S's argumentative theory, but perhaps we cannot rule out having done so in order to confirm a pre-existing hypothesis.

## Spontaneous inferences provide intuitive beliefs on which reasoning proper depends

doi:10.1017/S0140525X10002803

James S. Uleman, Laura M. Kressel, and SoYon Rim

Department of Psychology, New York University, New York, NY 10003.

jim.uleman@nyu.edu lmk323@nyu.edu soyon.rim@nyu.edu

<http://www.psych.nyu.edu/uleman>

<https://files.nyu.edu/lmk323/public/>

**Abstract:** *Spontaneous inferences* are unconscious, automatic, and apparently ubiquitous. Research has documented their variety (particularly in the social domain) and impact on memory and judgment. They are good candidates for Mercier and Sperber's (M&S's) "intuitive beliefs." Forming spontaneous inferences is highly context sensitive, varying with the perceiver's conscious and unconscious goals, and implicit and explicit theories about the domain in question.

Persuasive as the target article is in arguing that "reasoning proper" is implicitly intended "to devise and evaluate arguments intended to persuade" (abstract of the target article), it says too little about the unconscious "process of inference" that generates the "intuitive beliefs" that are input to this reasoning. This is a serious omission, because one cannot document how reasoning might select and shape arguments without specifying what the inputs to reasoning are. Recent work on spontaneous social inferences (e.g., Uleman et al. 2008) illustrates some of the methods and findings that may fill in this gap.

Spontaneous inferences are unintended, unconscious, practically effortless, typically uncontrollable, and apparently ubiquitous. Most research has been on spontaneous trait inferences (STIs; for an early review, see Uleman et al. 1996). Consider "John returned the wallet with all the money in it." When asked to memorize or merely familiarize themselves with such sentences, most people infer that John is honest. This has been shown with cued recall, lexical decisions, probe reaction times, savings-in-relearning, and false-recognition paradigms. People are more likely to assert that "honest" was in the sentence paired with John's photo than the sentence paired with Harry's photo, even though it was not (false recognition; Todorov & Uleman 2002). When people are subsequently asked to learn word pairs such as "John – honest," they do so more readily than "Harry – honest," even though they no longer recognize

which trait-implying sentence described John (savings-in-relearning). And they rate John (in a photo) as more honest, even though they cannot remember what he did or that they made any inference at all (Carlston & Skowronski 2005). So, as Mercier & Sperber (M&S) claim, these unconscious inferences provide the raw material for conscious judgments and presumably for the "reasoning proper" that justifies these judgments.

Spontaneous social inferences are not restricted to traits. There is good evidence that goals and situational (not just trait) causes of behavior are spontaneously inferred. When people read about unjust situations, they spontaneously activate such concepts as "unfair" and "injustice," but only when they imagine themselves being treated unfairly (Ham & Van den Bos 2008). They spontaneously infer causes of largely nonsocial events (Hassin et al. 2002). In these studies, the texts (or pictures; see Fiedler et al. 2005) are pretested by asking people for their conscious inferences. Stimuli that reliably imply whatever is of interest are then tested for spontaneous inferences. The same methods have been used to demonstrate that there are cultural and personality differences in who makes which inferences (see Uleman et al. 2008).

Multiple spontaneous inferences can occur simultaneously to the same stimuli. For example, Ham and Vonk (2003) showed that both dispositional and situational inferences during comprehension of a single event ("She got an A on the chemistry exam." → smart, and → easy). This suggests that, just as Swinney (1979) found that homonyms ("bank") initially activate multiple meanings ("money," "river") during text comprehension, multiple inferences occur spontaneously during the observation of events, and later selection among them occurs on the basis of wider contexts.

Like many concepts, traits have multiple meanings and uses (Uleman 2005). Traits can function as causal explanations of behavior, or traits can function as simple descriptions of behavior. The same is likely true of other concepts that are activated spontaneously. In explicit dialogue, the pragmatic context in which traits appear allows us to determine their intended meaning and function. But when inferences are spontaneous (i.e., unconscious), no such context exists. Recent research has shown that isolated trait terms function cognitively as causes, not merely as descriptions (Kressel & Uleman 2010). And subsequent unpublished work (Kressel 2010) shows that people with stronger implicit (as well as explicit) causal theories of traits' meaning are more likely to make STIs.

Such trait inferences can become associated with the "wrong" actors, in spontaneous trait transference (STT). If person A says something that implies a trait about person B, and only person A is present (or pictured), that trait becomes associated with person A (Skowronski et al. 1998). This does not occur if person B is also pictured, however (Goren & Todorov 2009). This suggests that spontaneously inferred concepts are easily "bound" to incorrect sources. Thus, events can spontaneously activate a variety of unconscious concepts and associations, all of which provide grist for the "reasoning proper" mill.

Which concepts are activated, and which associations or bindings occur, are context sensitive in other ways. Rim et al. (2009) have shown that, consistent with construal level theory, STI is more likely if the actor is more psychologically distant, either temporally or spatially. People think of distant things more abstractly, and traits are an important kind of social abstraction. Furthermore, unpublished data (Rim et al. 2010) show that nonconsciously primed goals can shape which inferences occur spontaneously and are bound to actors. Thus, nonconscious goals affect spontaneous inferences in several ways, all outside of awareness.

Finally, research on the logical inferences made during text comprehension goes well beyond bridging and predictive inferences. Lea (1995) showed that deductions according to *modus ponens* (if *p*, then *q*; *p*; therefore *q*) occur spontaneously, and Champion (2006) uncovered ways that certain and hypothetical

inferences differ. Thus, spontaneous inferences are not limited to the social domain. When stimuli present enough clear information and constraints, both logical and illogical inferences occur (e.g., Rader & Sloutsky 2002).

The formation of “intuitive beliefs” is more complex than the target article describes. Research on spontaneous inferences (social and otherwise) can tell us much about how intuitive beliefs are formed and what they are, before reasoning proper shapes them into persuasive arguments.

Incidentally, the argument that people can distinguish good arguments from bad, based on Petty and Cacioppo’s (1979) persuasion research, is completely circular. They have no principled basis for constructing good versus poor arguments; the arguments are simply pretested to have these properties.

## Query theory: Knowing what we want by arguing with ourselves

doi:10.1017/S0140525X10002797

Elke U. Weber and Eric J. Johnson

Center for Decision Sciences, Columbia University, New York, NY 10027.

euw2@columbia.edu ejj3@columbia.edu

**Abstract:** Mercier and Sperber (M&S) argue that reasoning is social and argumentative, and that this explains many apparently irrational judgment phenomena. We look at the relationship between interpersonal and intrapersonal argumentation and discuss parallels and differences from the perspective of *query theory*, a memory-based model of constructive preferences. We suggest an important goal is to integrate models across inference and preference.

Mercier and Sperber’s (M&S’s) provocative perspective suggests that inference has adapted to a social world where argumentation is common, and that many phenomena identified as reasoning errors are not errors but adaptive when considered as inferences embedded in a social world.

We agree that inferences are often constructed when confronting a problem, and that this construction is influenced by context rather than generated by unvarying algorithms. We suggest, however, that inference construction is affected not just by social forces but also by properties of human memory retrieval and multiple goals of decision makers.

This commentary describes parallels and differences between M&S’s argumentative hypothesis and a memory-based account of preference construction: *query theory* (QT). M&S cite two applications of QT as examples of reason-based choice with resulting choice inconsistencies; namely, the endowment effect (Johnson et al. 2007) and greater discounting of time during delay than during acceleration decisions (Weber et al. 2007).

However, QT is more than another example of reason-based choice. It provides *evidence* and *process-level specification* of the implicit memory-retrieval and argument-integration processes people use to evaluate choice options and demonstrates their causal role in arriving at a decision. Just as M&S unpack intuitive inference, QT treats intuitive preferences neither as a primitive (as in economics [Becker & Stigler 1977]) nor as a mysterious black box (as in social cognition [Dijksterhuis et al. 2006a]), but instead documents the cognitive mechanisms used in constructing preferences (Weber & Johnson 2006).

These are the key process specifications of QT: (1) People query past experience for evidence supporting different choice options, (2) these queries are executed sequentially, and (3) the first query produces richer representations because of output interference. This occurs because, as evidence for the first option is generated, evidence supporting other choice options is temporarily suppressed. Finally, (4) choice follows from the resulting balance of evidence. Since the order of options

consideration influences the balance of evidence, it is important to know what determines which choice option gets queried first. Characteristics of the choice environment often determine what option is considered first, such as the existence of decision defaults. Like M&S, QT suggests that framing effects occur because different frames make reasons differentially available. QT finds that framing works by influencing the order in which two options are considered and thus the balance of evidence, which mediates choice (Johnson et al. 2007; Weber et al. 2007). For example, different countries have different defaults for organ donation, which changes the order in which queries pro versus con donating are considered, producing different levels of organ donation (Johnson & Goldstein 2003). Similarly the order of consideration can be affected by different attribute labels that trigger positive versus negative emotions for different choice options (Hardisty et al. 2010).

Just like the inferential processes described by M&S, QT processes operate automatically, without awareness, and are effective (though sometimes biased) products of motivated cognition. The motivation for which option to consider first makes sense most of the time. Default options currently in place were typically selected for good reasons and have not caused injury or harm. Options that trigger desire have desirable features, and options that don’t trigger disgust or contempt are typically superior to those that do. Giving such options an advantage by querying arguments for their selection first is a way of making the right decision faster and with greater confidence. Both inference and preference trade off between accuracy and efficiency and confidence, though these different goals do not always work in opposite directions. Whereas argumentative goals raise confidence in one’s inferences or decisions and also shorten time to reach them, the initially favored options or opinions typically have good reason behind them, and the seemingly biased inference or preference reflects mostly reasonable Bayesian priors, with perhaps some built-in conservatism.

These parallels between M&S’s hypothesis and QT suggest that the purpose of argumentation is not purely interpersonal, but that implicit argument recruitment, in some outcome-biasing fashion, is also an intrapsychic process that is part of implicit preference construction. Note that Franklin’s comment about the human ability to “find or make a reason for everything one has a mind to do” (cited by M&S in support of social argumentation [sect. 4.1.4, para. 1]) was prompted by his internal struggle between vegetarian beliefs and the tempting smell of freshly caught fish on a sea voyage (Franklin 1817/2006). (He justified eating the fish by recalling the observation, while watching the fish being cleaned, that it had eaten other, smaller fish.) While this is an example of conscious inference, justification, and argumentation, M&S and QT argue that such memory retrievals and inferences occur constantly and without conscious awareness to guide our actions. Few choices offer dominating alternatives, and internal conflict between competing objectives and hence choice alternatives is the norm. Like Franklin’s example, many decisions also have the potential for postdecisional regret, making it important to bolster confidence that the chosen option is in fact the best one.

Are there differences between intrapsychic and interpersonal argumentation? One difference relates to one of the most crucial components of QT’s sequential argumentation; namely, the process of output interference, where the first query for evidence supporting the implicitly favored, and thus first-considered, choice option temporarily inhibits arguments for other choice options, hence reducing generation of support for it during subsequent queries. This is clearly an intrapsychic process, not an interpersonal one. It is only when I generate arguments for a given action (and not when someone else does so) that the accessibility of arguments for other actions in my mind gets inhibited. To the extent that we find similar subtle biasing of choices in line with the goals of motivated cognition in group discussion and decision settings, it must be accomplished

by different processes. Smith et al. (2009) suggest that interpersonal argumentation in group decisions changes goal-specific weights and not the order by which evidence is considered and thus its recall success. M&S find similar differences in implicit inference processes when inferences are made by a single individual or by several individuals in a group setting.

In summary, preference and inference tasks seem connected, among other things, by a common argumentative nature, which may suggest shared cognitive mechanisms (Weber & Johnson 2009).

## Reasoning, robots, and navigation: Dual roles for deductive and abductive reasoning

doi:10.1017/S0140525X10002955

Janet Wiles

School of Information Technology & Electrical Engineering, University of Queensland, Brisbane 4072, Australia.

wiles@itee.uq.edu.au <http://www.itee.uq.edu.au/~janetw/>

**Abstract:** Mercier & Sperber (M&S) argue for their argumentative theory in terms of communicative abilities. Insights can be gained by extending the discussion beyond human reasoning to rodent and robot navigation. The selection of arguments and conclusions that are mutually reinforcing can be cast as a form of abductive reasoning that I argue underlies the construction of cognitive maps in navigation tasks.

Mercier and Sperber's (M&S's) theory of the adaptive value of argumentative reasoning is intriguing from a computational perspective, since the search for arguments that support a given conclusion is computationally more difficult (viewed as a reasoning problem) than logical reasoning. The first logical solvers were developed in the 1950s (Newell & Simon 1956). Argumentative computers are yet to be developed.

*Argumentative reasoning*, defined broadly as the discovery of statements to support a given conclusion can be cast as a form of abductive reasoning, or inferring a precondition from a consequent (following Peirce 1931–35). Such reasoning is logically fallacious, but as M&S's target article details, it is typical of human behaviour to select arguments and conclusions that together are mutually reinforcing.

We accept M&S's arguments for the adaptive value of argumentative reasoning as a communicative skill. However, just as questions have been raised in other fields about the evolution of the sophisticated communicative abilities of humans, we can also ask how an argumentative ability could have evolved. Many evolutionary adaptations are thought to be exaptations; that is, new uses for existing structures. Verbal argumentative reasoning obviously draws on linguistic ability, but it need not postdate it. We consider the possibility that cognitive abilities underlying argumentative reasoning may predate the evolution of language by exapting abductive abilities from other domains.

Reasoning is not the only domain where adaptive behaviour may utilise abductive reasoning. A much more ancient evolutionary ability, which humans share with other mammals, birds, reptiles, and even insects, is the ability to navigate. Much is known about the navigational systems of mammals, including the neural representations of places (O'Keefe & Dostrovsky 1971) linked into cognitive maps (O'Keefe & Nadel 1978; Tolman 1948), grid cells (Moser et al. 2008), and head-direction cells (Taube et al. 1990). Complementing neural studies are computational models and embodied robots, and it is the fully functional robotic systems (Arleo & Gerstner 2000; Kuipers 2000; Milford & Wyeth 2003; Thrun 2003) that provide insight for this commentary.

Two approaches can be contrasted for robotic navigational systems: a logically correct approach based on Bayesian

reasoning (analogous to deductive reasoning), and one based on a bio-inspired approach that exploits a form of abductive reasoning to construct a cognitive map. In mobile robots, a key problem is to maintain an estimate of one's current location while exploring and mapping a new environment (called *simultaneous localisation and mapping* [SLAM]). Given information about localisation (such as a Global Positioning System [GPS]), mapping is a relatively straightforward deductive reasoning problem, and conversely, given a map, localisation is straightforward. However, when both tasks must be solved simultaneously (in the absence of GPS), the errors in each compound. Many locations do not have unique landmarks; apparently unique features of one environment may turn out to be present only transiently or to be shared by other locations. Even recognising a previously visited location at a later time can be challenging. In vision-only SLAM, one of the best-performing systems is the RatSLAM system (Milford 2008), inspired by the hippocampal mapping system of the rodent. Initially developed using place cells and head-direction cells, it was discovered early on that the robots also needed something akin to grid cells (although when the model was first developed in 2003, grid cells themselves were yet to be discovered). RatSLAM learns the paths that a robot traverses through its environment and links them into maps. It uses a unique optimisation system that maintains information that is locally consistent, while also estimating a global map.

If a location is considered a "conclusion" in a mapping task, and features of the environment are considered "arguments to support that conclusion," then systems that are effective at navigation are of necessity abductive reasoners. Maps are constructed by using locations for which there is evidence, and evidence is retained when it is useful for localisation. Maps and their evidence need to be mutually reinforcing to be useful. The hippocampus has been linked to many aspects of cognition as well as spatial memory. Argumentative reasoning may well be the latest of its exapted abilities.

## Some empirical qualifications to the arguments for an argumentative theory

doi:10.1017/S0140525X10002840

Christopher R. Wolfe

Department of Psychology, Miami University, Oxford, OH 45056.

WolfeCR@muohio.edu

<http://think.psy.muohio.edu/home/>

**Abstract:** The empirical research on the psychology of argumentation suggests that people are prone to fallacies and suboptimal performance in generating, comprehending, and evaluating arguments. Reasoning and argumentation are interrelated skills that use many of the same cognitive processes. The processes we use to convince others are also used to convince ourselves. Argumentation would be ineffective if we couldn't reason for ourselves.

Mercier and Sperber (M&S) are insightful in proposing a strong link between reasoning and argumentation. Understanding the argumentative context sheds light on the processes of reasoning. However, empirical research on the psychology of argumentation contradicts several of their key claims. Contrary to their position, reasoning biases are common even in the context of argumentation, the *confirmation bias* is not a feature of argumentation and actually weakens arguments, and people cling to claims less rigidly than is tacitly assumed by the authors.

M&S's review of the literature on the psychology of argumentation is surprisingly sparse. Unfortunately, the data suggest that people are subject to fallacies and suboptimal performance in generating, comprehending, and evaluating arguments. Kuhn

(2001) found that 4-year-old children are able to distinguish between evidence and explanation. However, children are often poor at generating arguments (Means & Voss 1996). In adolescents, national educational assessments find that only about 15% of 12th graders are adequately skilled in developing written arguments (Beatty 1996; Greenwald et al. 1999). Anne Britt and colleagues conducted a series of experiments on the comprehension, evaluation, and generation of argumentative texts by college students. Among their findings, only about 35% of participants were able to identify the main claims and reasons in arguments, 37% failed to reject unsupported arguments, 32% failed to reject unwarranted arguments, 48% included other-side information in their arguments, and 65% wrote essays that did not include a single reason to support their claims (Britt & Kurby 2005; Britt & Larson 2003; Britt et al. 2005; Larson et al. 2004). Britt et al. (2008) found that university students have difficulty precisely recalling the main predicate of argument claims, but less difficulty recalling the predicates of comparable narrative statements. Wolfe et al. (2007) found that highly implausible reasons and warrants (for example, Paul should walk to the store “because walking is the absolute best exercise that will ever exist, and exercising can lead to immortality”) yielded higher agreement than the same claims without support. Argumentation is a fundamental skill that permeates human thinking (Voss & Van Dyke 2001). However, M&S have painted an unduly optimistic portrait of our argumentation abilities.

Some of what M&S describe as “confirmation bias” is sometimes called the “myside bias” (Baron 1995; Perkins et al. 1991; Toplak & Stanovich 2003; Wolfe & Britt 2005; 2008; Wolfe et al. 2009a). Although some authors use the terms interchangeably, *confirmation bias* typically refers to a biased search for or weighing of evidence, whereas *myside bias* refers to biases in generating reasons or arguments (Wolfe & Britt 2008). M&S state that the confirmation bias “is a consequence of the function of reasoning and hence a *feature* of reasoning when used for the production of arguments” (sect. 3, para. 1, emphasis theirs). My colleagues and I have conducted a series of experiments on argumentation, and the evidence differs in key respects from their assertions (Wolfe & Britt 2005; 2008; Wolfe et al. 2009a).

Wolfe & Britt (2008) had participants write argumentative essays under different conditions. Some were assigned to write essays for and others against an unpopular proposition. Participants had access to a number of pro and con online texts, and we also examined their search behavior. We found that the myside bias was pervasive. However, it was not linked to participant’s personal opinions. People exhibited the myside bias when arguing for the side with which they personally disagreed just as often as for the side with which they agreed. We have replicated this finding, yet also find significant correlations between opinion strength and myside bias on nonargumentative reasoning tasks (Wolfe & Boone, under review; Wolfe & Britt 2008). Moreover, participants exhibiting the myside bias in their arguments were not biased in their search. They sought out both pro-side and con-side texts.

The myside bias is indeed a bias – even in the context of argumentation. To illustrate, a content analysis of published authentic arguments found that most writers included other-side information in their arguments, commonly for the purpose of rebuttal (see Wolfe & Britt 2008). In laboratory experiments, presenting and rebutting other-side information consistently leads to better ratings of agreement, argument quality, and impressions of authors than does excluding other-side information (Wolfe & Boone, under review; Wolfe et al. 2009a). The myside bias weakens arguments measurably.

The factors predicting the myside bias in written essays are individual differences in beliefs about argumentation. Evidence stems from successful tutorials that significantly reduce the myside bias in generating and evaluating arguments (Wolfe et al. 2009a; 2009b), answers to the open-ended question “what

makes a good argument?” (Wolfe & Britt 2008), and reliable individual difference measures (Wolfe & Boone, under review; Wolfe & Britt 2009). The context of argumentation changes the nature of the myside bias, but one-sided argumentation is problematic and not an inherent feature of argumentation.

A tacit assumption in M&S’s account is that people have strong, stable preferences and unwavering commitments to claims. Argumentation is seen as a form of rationalization used to convince others of claims derived from intuitive processes about which people are only dimly aware. Yet, starting with early research on informal reasoning (Perkins et al. 1983), we have learned that positions are often fluid and tentative. As reasoning progresses, those positions undergo changes. We typically argue about matters that are “debatable,” where reasonable people arrive at different conclusions and are open to persuasion. I believe that reasoning and argumentation are interrelated skills drawing on many of the same cognitive processes – two sides of the same coin. Dual-process theories suggest that people lack access to many of our own cognitive processes. Thus, when we have tentative intuitions that are not well understood, we use many of the same processes to convince ourselves that in other contexts we use to convince other people. However flawed these processes may be, argumentation would be ineffective if we were not also able to reason for ourselves.

## Deliberative democracy and epistemic humility

doi:10.1017/S0140525X10002888

Kevin Chien-Chang Wu

National Taiwan University College of Medicine, Taipei, Taiwan.  
ccwu88@ntu.edu.tw

**Abstract:** Deliberative democracy is one of the best designs that could facilitate good public policy decision making and bring about epistemic good based on Mercier and Sperber’s (M&S’s) theory of reasoning. However, three conditions are necessary: (1) an ethic of individual epistemic humility, (2) a pragmatic deflationist definition of truth, and (3) a microscopic framing power analysis during group reasoning.

In recent decades, we have witnessed many public policy scandals and controversies in which, compared with laypersons, experts are not necessarily better in either epistemic or moral aspects (Godlee 2010; Jasanoff 2003; Wynne 1996). Following the loss of public trust in expert decision making, new discourses of deliberative democracy for public policy making have appeared (Lengwiler 2008). Examples in the new trend of emphasizing public participation are consensus conferences, citizens’ juries, citizens’ panels, and the like (Abelson et al. 2003). Usually, there are two meanings of *deliberation*: The first is “a particular sort of discussion” that “involves the careful and serious weighing of reasons for and against some proposition” (Fearon 1998, p. 63). The second is “an interior process” of reason weighing by an individual (Fearon 1998, p. 63). Deliberative democracy adopts the first meaning, so deliberative democracy fits in Mercier and Sperber’s (M&S’s) group-reasoning situation.

Many theoretical arguments support deliberative democracy, but here I focus on the epistemic aspect of deliberative democracy. According to John Dewey’s experimentalist account of democracy, the epistemic power of democracy is inseparable from citizens’ collective deliberation. In a democratic regime, citizens are engaged to deliberate collectively over the foreseen consequences of public policies and then choose, test, evaluate, and revise the policies implemented. Accordingly, the discussions among citizens with diverse backgrounds and local knowledge could facilitate the fair recognition of public interest

problems and the adoption of public policy proposals that are comprehensive enough. Public policy choice and implementation is like an experiment in whether dynamic feedback from the policy implementation will render another session of democratic deliberation (Anderson 2006). Also, according to another pragmatist, C. S. Peirce, a proposition could be true if it can survive the test of best reasons, evidence, and arguments (Misak 2009). Since almost no one would deem the seeking of truth an unworthy goal, it holds that deliberation in a democratic regime would offer the best chance of achieving this goal (Talisie 2009).

As already mentioned, the typically adopted concept criteria of deliberative democracy include the state of disagreement among the participants and the task of reaching collective decisions (Thompson 2008). Therefore, a good design of deliberative democracy would be compatible with the propositions by M&S about reasoning for arguments. Accordingly, all individuals, no matter whether they are experts or laypersons, are subject to proactive reasoning. When these individuals deliberate alone, they usually reason to confirm, rather than scrutinize, their original arguments. It is better to conduct group reasoning such that each member can contribute to mutual non-proactive evaluations of arguments and conclusions not in their favor. For M&S, “[a]rgumentation is uniquely effective in overcoming disagreements that are likely to occur, in particular in relatively egalitarian groups” (sect. 1.2, para. 9).

Also, to legitimize the conclusions reached through deliberative democracy, it is usually proposed that in the process of mutual justification (presenting and responding to reasons intended to justify a political decision [Gutmann & Thompson 2004]), deliberative democracy should have such characteristics as public spiritedness (arguments for common good), equal respect for each participant, accommodation (retaining the possibility of collaboration on other issues), and equal participation (no domination phenomenon) (Thompson 2008). The epistemic good of deliberative democracy comes from engaging people with different motivations for group reasoning and subjecting each version of reasoning to mutual scrutiny. Thus, current theoretical proposals have shown that deliberative democracy is one of the best designs for facilitating good public policy, as implied in M&S’s theory.

Empirical research into the effectiveness of deliberative democracy is still in its burgeoning stage, but the qualified promise of deliberative democracy demonstrates that deliberation is not an easy task (Ryfe 2005). A good design of deliberative democracy should establish rules to maintain the theoretical aims, allow people to tell stories to make cultural meanings in addition to making cognitive sense, encourage leadership that facilitates deliberation, endeavor to relate the outcome of deliberation to the participants, and finally create environments to facilitate learning how to deliberate (Ryfe 2005). All these complicated issues were not addressed by M&S and could supplement their propositions as applied outside of the experimental fields.

Three conditions are needed for deliberative democracy to achieve epistemic good. First, when it comes to equal participation in deliberative democracy, the explicit and implicit standards for the concepts used in the communication might lead to the exclusion of the concepts used by the marginalized groups to make cultural meanings. In this kind of “hermeneutic injustice” (Fricker 2007), the majority’s conceptual framework might squeeze out or mask the minority’s specific experiences. For example, in John Rawls’ formulation of public reason, if the contents of claims contain concepts that are not easily available and assessable by the public, these claims would be excluded from the public domains because they contain no recognition-worthy reasons (Morgan-Olsen 2010). However, we should not forget that the frames and ways we observe and analyze things often constrain our decisions and results (Wu 2008). Therefore, we should uphold an ethic of epistemic humility by which we

take serious novel concepts and experiences presented in deliberation.

Second, we should recognize that if the epistemic goal of deliberative democracy were to seek truth, then the dynamism in the process would only point us to the deflationist version of truth. There is no way we could ascertain the realist version of truth through deliberation over the feedback from the implementation of public policies. Sticking to the realist version of truth would disrupt the function of deliberation, bringing it to a halt. Third, pursuing the previous two conditions, we have to analyze the explicit and implicit operations of microscopic framing powers empirically during group reasoning. Here, I do not insist on the abolition of all these operations of framing powers (Lengwiler 2008), but we should understand whether the minority concepts and frames are given their due course for sense making and meaning making. Based on the findings, we could further design the rules that would meet the requirements of epistemic humility.

## Authors’ Response

### Argumentation: Its adaptiveness and efficacy

doi:10.1017/S0140525X10003031

Hugo Mercier<sup>a</sup> and Dan Sperber<sup>b</sup>

<sup>a</sup>Philosophy, Politics and Economics Program, University of Pennsylvania, Philadelphia, PA 19104; <sup>b</sup>Jean Nicod Institute (EHESS-ENS-CNRS), 75005 Paris, France; and Department of Philosophy, Central European University, Budapest, Hungary.

hmercier@sas.upenn.edu dan@sperber.fr  
http://sites.google.com/site/hugomercier/ http://www.dan.sperber.fr

**Abstract:** Having defended the usefulness of our definition of reasoning, we stress that reasoning is not only for convincing but also for evaluating arguments, and that as such it has an epistemic function. We defend the evidence supporting the theory against several challenges: People are good informal arguers, they reason better in groups, and they have a confirmation bias. Finally, we consider possible extensions, first in terms of process-level theories of reasoning, and second in the effects of reasoning outside the lab.

We are grateful to the commentators for their support, their insightful criticisms, and their useful suggestions. Even when we felt we were being misunderstood, we learned – or so we hope – how to make our case in a clearer and more compelling manner. In answering these commentaries, we focus in turn on the evolution and the function of reasoning, on its strengths and biases, and on its mode of operation. But first we have to defend our definition of reasoning against several challenges.

### R1. Different definitions of reasoning

In the target article, we defined reasoning as the mental act of constructing or evaluating an argument that is at least partly explicit. In particular, it must contain both a conclusion and reasons to accept this conclusion, even if some of the steps leading from these reasons to the conclusions are left implicit. In this sense, reasoning is in

contrast with ordinary intuitive inference, a process that yields a conclusion without articulating the reasons to accept it. So far, our definition is close enough to philosophical and commonsense use of the term *reasoning* and at odds with the now widespread use in psychology of “reasoning” as a mere synonym of *inference*. Needless to say, several definitions of reasoning may each target a phenomenon worth studying.

In line with much evolutionary psychology, we assumed that the mind is an articulation of many fairly specialized mechanisms; in particular, mechanisms of intuitive inference that are specialized for a specific cognitive domain or task such as recognizing faces, interpreting their expressions, processing frequencies, and decoding speech. Even though it contrasts with ordinary intuitive inference, reasoning as we describe it is itself a form of higher-order intuitive inference with a specialized domain and task: It delivers intuitions about reasons-conclusions relationships. This way of distinguishing reasoning proper from other inferential mechanisms is largely rooted in the same kind of observations and concerns that have led to the development of *dual-process* or *dual-system* approaches to reasoning (as we have discussed in greater detail in Mercier & Sperber 2009).

Several commentaries defend a different definition of reasoning that may be closer to that of inference in general or to a more standard dual-process approach to system 2 reasoning.

**Khlentzos & Stevenson** suggest that some type of system 2 reasoning must have evolved to arbitrate between contradictory system 1 outputs. Unless a very liberal definition of system 2 is adopted – one that encompasses much more than reasoning as it is defined here – it seems as though such problems are routinely solved by system 1 itself. Any cognitively complex organism will be faced with contradictory outputs – for instance, when perception contradicts memory because the environment has changed – and must have ways to arbitrate among them. Reasoning may help accomplish this task in some difficult cases, but it is specifically geared toward this end – and this would be true even with a much broader definition of the term *reasoning*.

**Poletiek** makes a comparable proposal regarding action selection, using the example of people looking for their keys and having to choose between different search strategies. According to her, it is mostly a consideration of costs and benefits that decide whether the search is “confirmatory” or “falsificatory.” But, as we suggested, such choices are commonly made without reasoning proper (a point Poletiek seems to agree with). Evolved intuitive mechanisms of action selection are designed to take into account costs and benefits without producing the reasons for their choices. Reasoning, on the other hand, produces such reasons and hence a justification for a course of action without directly triggering it. **Narvaez** suggests in the same vein that reasoning “includes figuring out what course of action to take” (para. 2). While reasoning is sometimes used in decision making – as reviewed in section 5 of the target article – the vast majority of our actions are guided by intuitions and so fall outside of the scope of reasoning as defined and thus outside the scope of the article, as well.

More generally, thinking and action selection involve more than just domain-specific intuitive inferences and

reflective reasoning. In many dual-process models, system 2 is in fact likely to encompass mechanisms other than reasoning. **Evans** suggests hypothetical thinking – the ability to represent future situations. **Narvaez** refers to some forms of elaborated planning. **Connolly & Reb** talk of mechanisms designed to avoid decisions we would regret, making interesting suggestions regarding ways to eliminate some reasoning biases – in particular reason-based choice. These authors point out that these mechanisms can directly lead to good outcomes without involving argumentation, and see this as an objection to our evolutionary argument. But these mechanisms do not qualify as reasoning under our definition – they are not necessarily reflective, they do not deal with arguments, and so on. Still, these suggestions point towards an interesting direction of research. While system 1 is commonly seen as a set of difference mechanisms, system 2 is often considered to be more unitary. It is also possible however to view system 2 as comprising several different mechanisms, such as reasoning, planning, imagination, and strategic thinking, each with a specific function. What might justify seeing these different mechanisms as part of a single system is, for instance, their heavy use of working memory or of metarepresentational machinery. If different system 2 mechanisms shared such common resources, this might help explain the covariation of traits measured by various measures of cognitive ability stressed by Evans.

Our definition of reasoning may be debatable, but the argumentative approach to reasoning is about reasoning as we defined it. To object to this definition, it is not enough to offer another definition that may be reasonable and useful. What would have to be shown is that ours fails to identify a phenomenon with enough autonomy and integrity to be a proper object of study and insight.

## R2. Evolution and function of reasoning

A number of objections and suggestions were based, we feel, on a partial or, in some cases, mistaken understanding of our hypothesis on the evolution and the function of reasoning. The misunderstanding we are most eager to correct consists in attributing to us the view that reasoning has only rhetorical rather than both rhetorical *and* epistemic goals. We didn’t argue that reasoning is designed only to find arguments in order to persuade others (**Godfrey-Smith & Yegnashankaran; Poletiek**). We don’t hold that epistemic goals should be poorly served by reasoning (**Khlentzos & Stevenson; Kuhn**), or that mere rhetoric is all it takes to influence people (**Narvaez; Sternberg**). Nor does it follow from our account that people should hardly ever change their mind (**Wolfe**). On the contrary, reasoning evolved in part to make people change their mind by giving them *good* reasons to do so. These misunderstandings may be linked to the fact that, in the target article, we devoted more space to the production of arguments by communicators (whose goal is indeed to persuade) than to the evaluation of these arguments by the audience (whose goal is to be informed). This imbalance reflected the present state of the literature we surveyed rather than a theoretical bias. Actually, the argumentative theory would not make evolutionary sense if arguments



were addressed to people who were wholly unable to evaluate them from a sound epistemic perspective.

### **R2.1. The double-sided argumentative function of reasoning**

Why does reasoning exist at all, given that it is a relatively high-cost mental activity with a relatively high failure rate? To answer this question, we proposed to step back from the study of individual cognitive processes and to look at the evolution of human communication. Humans are immersed in a flow of socially transmitted information and are highly dependent on it. For communication to have evolved, it had to be advantageous to both communicators and receivers (who are, of course, the same individuals but acting in two different capacities). What makes communication advantageous to receivers is that it provides them with rich information that they could not, or not easily, have obtained on their own. For this, the information they receive has to be genuine information; that is, close enough to truth. What makes communication advantageous to communicators is that it allows them to achieve some desirable effect in the receivers. For this, the information they emit has to be conducive to this effect, whether it is true or false.

**Dessalles**, who has himself developed a perspective in some respects comparable to ours, understands us to claim that “the biological function of reasoning is to achieve *shared knowledge optimization*” (para. 2) and that this is done not at the individual but at the group level. We do argue that the main function of reasoning is indeed social but by serving the social interests of individuals rather than the collective interests of the group.

To reap the benefits of communication while limiting the risk of being misled, receivers must exercise what we have called *epistemic vigilance* (Sperber et al. 2010). There is no fail-safe algorithm to sort genuine from spurious information; hence, we argue, various cost-effective heuristics that may contribute to approximating such a sorting are likely to have evolved. The main heuristic of epistemic vigilance consists in assessing the trustworthiness of communicators. Thus, we agree with **Opfer & Sloutsky** that “children detect trustworthiness long before they detect argument inconsistency” (para. 4) (e.g., see Mascaro & Sperber 2009). But if detecting the trustworthiness of communicators were the only heuristic used, then receivers would end up rejecting a good amount of genuine and relevant information when they lack sufficient ground to accept it on trust. For instance, few if any readers of our article would accept its conclusions just out of trust in its authors! To be more effective, epistemic vigilance must be exercised not only towards the source of information but also towards its content. Independently of its source, a message may have a greater or a lesser believability. This believability is assessed by considering its coherence with background knowledge. *Coherence checking*, we argue, is the second major heuristic used in filtering communicated information, and is at the basis of reasoning proper.

Coherence checking starts as a method for receivers to filter information; it ends up being exploited also by communicators who engage in coherence displays in order to have their messages accepted. Just as receivers would filter out some genuine information if they relied only on

the trustworthiness of the source, communicators would fail to communicate some believable messages if they relied only on their own authority. Arguing consists in displaying coherence-based reasons for the acceptance of a given message. It is, in essence, an “honest display” strategy opened to evaluation and aimed at the audience’s epistemic concerns. Of course, what is displayed may be an invalid argument made in the service of deception and in the hope that its invalidity won’t be detected. Contrary to what **Dessalles** attributes to us, we do not believe that communicators argue in order “to correct or update others’ beliefs” (para. 3) when it is not to their advantage. They argue for whatever it is advantageous to them to have their audience believe. Often enough, for instance, when communicating to coordinate action, communicator and audience have convergent interests in sharing true information, but this is far from being always the case.

In an evolutionary perspective, receivers’ coherence checking creates selective pressure for communicators’ coherence displays in the form of arguments, which in turn creates selective pressure for adequate evaluation of arguments on the part of receivers. At least in some cultural contexts, this results in a kind of arms race towards greater sophistication in the production and evaluation of arguments. Of course, argumentation can be misused and abused – for instance, by arguing above the head of one’s audience (Sperber 2009) or by lacing arguments with appeals to emotion. Doing so, however, is more likely to serve the interests of the communicator than those of the audience. Contrary to what **Opfer & Sloutsky** maintain, “hot” persuasion is advantageous to communicators only to the extent that receivers yield to it, but it is not advantageous to receivers who care to be well informed. For this they had better reason, as Petty and Cacioppo (whom Opfer & Sloutsky oddly cite in support of their claim) have shown in numerous experiments that demonstrate precisely this point: When people are motivated to reason, they do a better job at accepting only sound arguments, which is quite generally to their advantage (e.g., see Petty et al. 1981).

### **R2.2. Other functions of reasoning?**

Several commentators, while agreeing that argumentation may be an important function of reasoning, suggest that it may serve other functions, as well: either social functions other than the production and evaluation of arguments (**Baumeister, Masicampo, & DeWall [Baumeister et al.]; Dessalles; Frankish; Pietraszewski**) or functions contributing to individual cognition (**Evans; Frankish; Godfrey-Smith & Yegnashankaran**). We recognize the possibility. Our claim is that argumentation is the *main* function of reasoning, and we tried to demonstrate this claim by showing the signature of this function in the way reasoning actually functions, and in particular in what has been seen as flaws and that we argue are features of reasoning. Any evolved mechanism can be put to a variety of uses: The mouth can be used to play a wind instrument, the skin can be used to display tattoos, and the sense of smell can be used to develop wine expertise. Whether these additional uses become functions in the biological sense, or even main functions (in which case Gould and Vrba’s term *exaptation* might be properly applied) depends on the relative

contribution these uses make to the fitness of organisms endowed with the mechanism.

**Dessalles** and **Frankish** suggest that argumentation could have evolved as a means to display one's intellectual skills. Indeed, argumentation can be put to such a use. However, the human drive to show off is so strong that just about everything in our physical and cognitive makeup has been recruited to that effect. People may indeed run, fight, or argue to impress others. But what makes these performances not only impressive but relevant is that running, fighting and arguing are useful for purposes other than impressing others. Moreover, the production of arguments typically involves a high degree of mere *satisficing* (Mercier, submitted a). People do not look for the best formulation of the best possible argument. Instead, they use the first minimally decent argument that comes to mind. If it works, then it wasn't worth expending more effort. If it doesn't, it is easy to try a rebuttal or another argument. This way of arguing is to be expected only if the goal is to convince, but it would be very surprising if reasoning had the function to display one's skills by producing impressive arguments. In other words, reasoning is more like a crow's than a peacock's tail: It may be a bit drab, but it serves its main function well. Its occasional use, for instance, in academic milieus, to display one's intellectual skills is unlikely to contribute to fitness to the point of having become a biological function, let alone the main function of reasoning.

**Pietraszewski** rightly points out that argumentation is used not just in the defense of factual claims but also of claims that are not properly speaking matters of fact but more matters of choice or of social alignment. He distinguishes two classes of cases, one where the goal of arguments is "to change representations of the payoff structure of pursuing certain future plans" (para. 8). When argumentation is used for such purpose, it recruits intuitions that bear on the domain at hand – for example, what is advantageous or what is well regarded – and puts forward reasons for a conclusion. Such cases may fall squarely under the characterization we have given of reasoning. Pietraszewski also draws attention to a second class of cases where "agreement and disagreement . . . become commodities in themselves as a way of signaling the coordination strength and challenging others. This class of argumentation psychology should be designed to conflate evaluations of the argument with the source and social context of the argument; who is arguing should be just as important as what they are saying when considering the 'goodness' of an argument" (para. 9). This is a welcome and important observation, but does it point to another function of reasoning? We would suggest rather that it highlights that communication typically involves a mix of means and goals. The main relevance of a communicative act may be in its explicit content, in its implicatures, or in the very fact that it took place at all (Sperber & Wilson 1995); it may have to do with transmission of factual information or, indeed, with signaling agreement and disagreement. This can be done in particular by using arguments not so much to convince but to polarize. The phenomenon of polarization that, in the target article, we discussed mostly in negative terms, should probably be seen as fulfilling a function along the lines suggested by Pietraszewski. This said, it is not clear that it bears so much on the function of reasoning rather than on a wider range of

mechanisms of social cognition that exploit and interact with argumentation in a variety of communicative interactions. In particular, as **Opfer & Sloutsky** insist and as Sell (2006, quoted by Pietraszewski) has shown with the example of anger, "hot" communication interferes with argumentation in important ways, reminding us that argumentation is only one device for persuasion among several.

**Baumeister et al.** draw attention to two major phenomena linked to reasoning: consciousness and culture. We noted that reasoning is a typically conscious activity but we did not elaborate for lack of sufficiently clear ideas regarding consciousness in general and its relationship to reasoning in particular. Baumeister et al. propose to extend the social hypothesis towards conscious thought in general: "Conscious thought enables people to talk to others and thereby enables small groups to resolve differences" (para. 6). Their arguments are indeed very congenial to the argumentative theory. Reasoning could then be seen as one part of a larger set of mental processes that are not typically thought of as being social by nature, even though they actually are. Baumeister and colleagues have made forays in this direction, and we hope that more research will follow, shedding new light on well-known but still puzzling results.

Regarding the role of reasoning in culture, we agree that indeed it is a source of cultural innovation (for example, in technology, law, and the sciences) and that it plays a role in cultural communication (with great cultural variations, we surmise). However, we are not convinced that this is *the function* or even *a function* of reasoning. Reasoning is advantageous to individuals who are better able to persuade others or to discriminate good from bad arguments. Most of the arguments humans produce and evaluate are about very local concerns: Who forgot to close the door? Should we walk or take the bus? Was John lying or mistaken? Reasoned answers to these local questions don't ever reach a cultural level of distribution. Reasoning in order to improve laws or to discover new techniques is very rare. It may well be beneficial to the group, but that is not enough to assume that reasoning evolved, through group selection, for such a benefit.

**Godfrey-Smith & Yegnashankaran**, drawing on Vygotsky, make the interesting suggestion that "reasoning is deliberative [i.e., individualistic] in function but dialogic in structure" (para. 3). We do not dispute that reasoning can be used for individual ratiocination. Even though solitary reasoning may not be the most effective way to enhance individual cognitive goals, we do not exclude that such enhancement may have sufficiently contributed to the relative fitness of reasoners to constitute a function of reasoning. What we would dispute is that this is at all likely to be the main function of reasoning, and the Godfrey-Smith & Yegnashankaran commentary is helpful to spell out the argument. If they are right and individualistic reasoning is dialogic in structure, why should this be so? Were a computer scientist to design a reasoning program (not specifically aimed at argumentation), would she opt for such a structure? In fact, many reasoning programs have been developed in artificial intelligence (AI) and in the psychology of reasoning (e.g., Johnson-Laird 2006; Rips 1994), and typically they generate arguments (in the formal sense of the term) without any dialogic back-and-forth between two agents. We suggest that this dialogic structure of individual reasoning is the

signature of its primarily argumentative main function. We would argue moreover that the main contribution of reasoning to individual cognition is in helping people evaluate other people's arguments.

Both **Evans** and **Frankish** suggest an interesting way to reconcile our suggestion that reasoning may have evolved for argumentation and the more classical view that it serves first and foremost individual cognitive goals. This individualistic function, Evans argues, might be an exaptation from an initial argumentative function favored by the concomitant evolution of "language, metarepresentation, and large forebrains" (para. 6). Frankish proposes that reasoning, even if it primarily evolved for argumentation, might have been substantially "co-opted to play a role in individual cognition" (para. 5). As we said in the preceding paragraph, we do not exclude the possibility that reasoning may also be adaptive in helping individual ratiocination. Still, our argument remains that the many apparent failings of reasoning – our knowledge of which owes much to Evans himself – make better sense as the signature of a main argumentative function.

**Evans** insists on the role of reasoning in anticipating the future. While anticipating the future is a major aspect of human cognition, it is not clear to us what specific role reasoning plays in this process, by contrast with other abilities such as imagination and simulation. It is even less clear what features of reasoning, if any, are specifically tailored for this role. **Frankish** points out that reasoning can be used to strengthen our resolve by buttressing our decisions with supporting arguments. Indeed, reasoning can do exactly that. However, if weakness of resolve had been enough of a problem to favor the evolution of a mechanism to deal with it, presumably the natural selection solution would have been to strengthen our resolve directly rather than to adapt reasoning to find resolution-strengthening arguments. Moreover, this very tendency to use reasoning to bolster our beliefs and decisions is likely to have more negative than positive consequences, as highlighted in section 4.2 of the target article.

### R3. Strengths and biases of reasoning and argumentation

#### R3.1. Are we really good at argumentation?

If reasoning evolved for argumentation, humans should possess decent argumentative skills: They should be able to engage in an informal debate, constructing, evaluating, and rebutting arguments. Several commentators question that this is so, pointing to research that shows significant improvement in argumentative ability through learning and casting doubt on the data we presented to defend argumentative skills.

**Harrell, Kuhn, and Wolfe** all point to very interesting data showing improvement in argumentation skills with training – as well as the relative poverty of these skills before training. Most of the studies cited, however, bear on *meta*-argumentative skills. Kuhn et al. (2008) report gains in "meta-level communications about the discourse" (p. 1310). Larson et al. (2004) study the understanding of long written arguments. Critical thinking skills are often evaluated through the ability to draw argument schemas or to recognize the different components of an argument and their relationship to one another. Such skills may very well be crucial for success in modern academic life

or even be of relevance to participation in democratic life, and their study is important in its own right. However, they are not part of what one can expect basic argumentative skills to be. Language provides a useful analogy. One can be a fluent language user without being able to tell what a verb is or, a fortiori, to parse a sentence. Likewise, one can be a skilled arguer without being able to recognize argument forms or draw argument schemas. In both cases, these abilities can help – a mastery of syntax can make one a better poet, a mastery of argument schemes a better essayist – but they are not necessary.

One study seems to tap into a more fundamental argumentative skill – the ability to draw *counterarguments* (Goldstein et al. 2009). In this experiment, sixth and seventh graders had to choose between a counterargument and a novel argument in a fictitious discussion, and they tended to choose the latter instead of the former. Even though the context successfully mimicked a debate, it does not follow that the participants were highly motivated to defend a position that had been arbitrarily assigned to them. In such a context, it is not surprising that they should engage in satisficing and pick an argument that might just be enough. Moreover, although the counterargument option might have been formally superior, it is not clear that it would have been the most efficient one in a real discussion.

**Harrell and Wolfe** also dispute the data used as evidence of people's basic argumentative skills. Harrell suggests deflationary interpretations for several of the findings we cite, but, however ingenious, these interpretations are not compelling. To be more specific: Bailenson and Rips (1996) do observe that being the first speaker increases the burden of proof. But with contentious first sentences such as "abortion should not be illegal" (p. S7), this is hardly surprising – or nonnormative. In Rips (2002), the repetitions are nothing but innocuous, and so participants are right to discount arguments accordingly. Neuman et al. (2006) do not really observe worse performance among people who argue. Participants were merely avoiding pointing out a fallacy in a polite dialogue (cf. Rips 2002, n. 1), but they were still perfectly able to detect it in a rougher discussion. Thompson et al. (2005) observed more belief bias when people were engaged in an argument ... because people were asked about their opinion and not logical validity. Far from having "difficulty producing evidence for a claim" (Harrell, para. 7), the participants in the study by Sá et al. (2005) were able to muster nearly six arguments to defend their opinion on difficult topics (for example, Why do children fail in school?). If these arguments were not always the best, this is only to be expected in such circumstances (as explained in sect. 2.3 of the target article).

**Uleman, Kressel, & Rim [Uleman et al.]** point out that the strong and weak arguments used in most persuasion research are "pretested ... to have these properties" (para. 10), so that it is hardly surprising that strong arguments should be more convincing. That researchers want to make sure their intuitions about argument strength are appropriate does not mean that they are without grounding, however. For instance, when trying to persuade students that a new exam would be beneficial, a strong argument might be that "average starting salaries are higher for graduates of schools with the exams" (Petty & Cacioppo 1979, p. 1921) accompanied by relevant statistics. By contrast, a weak argument might read "by not administering the exams, a tradition dating back to the

ancient Greeks was being violated.” It is not a stretch to find some objectivity in this judgment of strength: It is not a mere experimental artifact that the stronger arguments, in these experiments, do actually provide more support for their conclusion. Finally, **Wolfe** mentions a study in which “implausible reasons and warrants ... yielded higher agreement than the same claims without support” (para. 2). Famously, Langer et al. (1978) had already demonstrated the effect of poor reasons more than 30 years ago. However, they also showed that the effect mostly disappeared when the stakes increased – as argued in the target article, whether people genuinely evaluate arguments depends on how much they care about (and disagree with) the conclusion.

So, without denying that spontaneous argumentation skills are imperfect and can be improved by teaching – and that this is linked to the variable importance given to argumentation in different cultures and institutions – we maintain that they display a remarkable superiority to the reasoning skill elicited in nonargumentative contexts.

### R3.2. How efficient is group reasoning?

This question has elicited contrary opinions from the commentators. **Khleutzos & Stevenson** think that good performance in groups is obvious since “subjects share a common goal of finding the correct solution” (para. 11). However, the same participants can face the same problems with the same goal but individually fail, so having the correct answer as a goal can hardly be the whole story. **Johnson**, on the other hand, questions the generality of good group performance, and **Sternberg** claims that groups are actually very poor at reasoning or decision making, citing as a support the groupthink syndrome or group polarization. First, it should be stressed that the argumentative theory does not predict that groups will always make better decisions, but merely that reasoning should work better in the context of a genuine debate. Many other factors besides reasoning can impact the outcome of a discussion – strategic considerations, face saving, and so forth. And reasoning in group can also bring poor outcomes when there is no genuine deliberation. Actually, in section 2.3 of the target article, we offer an explanation based on the argumentative theory for group polarization. Without a refutation of this explanation, we don’t see how this very phenomenon can be used as evidence against the theory. Finally, **Opfer & Sloutsky** mention one study that showed groups of children performing more poorly after a discussion (Levin & Druyan 1993). It is true that sometimes the best arguments will point in the wrong direction. When, in the early 20th century, geologists argued against Alfred Wegener’s theory of continental drift, their mistaken conclusions stemmed not from poor reasoning but from the state of knowledge at the time. Moreover, the explanation offered by Opfer & Sloutsky – that the group member with the correct answer is simply more confident – has already been refuted in section 2.3 of the target article (for instance, how could that account for groups performing better than their best member?).

### R3.3. The strength of the confirmation bias

When we look for arguments in a debate, we are mostly interested in arguments for our side or against the other

side. This is why, we surmised, the *confirmation bias* is a feature of reasoning, at least in its role of argument producer. **De Neys, Poletiek, Stuppel & Ball**, and **Wolfe** have reservations about the prevalence and robustness of this bias.

**Poletiek** questions the evidence from hypothesis testing problems cited in support of the idea that the confirmation bias is specific to reasoning (sect. 3.1). We agree with Poletiek that most of hypothesis testing is actually not directed by reasoning, and that confirmatory strategies are the result of heuristics that do not display a genuine confirmation bias. But this does not explain why people fail to adopt falsificatory strategies when they are asked to and adopt them spontaneously when they test someone else’s hypothesis. It seems as though reasoning is unable to correct our own intuitions even though it can easily try to correct those of others.

**Wolfe** mentions a number of studies about the *myside bias*; that is, the tendency for participants to mostly – or only – give arguments that support their opinion. One of these studies show that, although “pervasive,” the *myside bias* could be oriented by the instructions (Wolfe & Britt 2008). Participants could be made to write essays against their opinion about a pretend requirement to “impose a 2-year math requirement for all students” (p. 8). But in this experiment, participants did not have to generate the arguments themselves. Instead, they were provided with a series of arguments for and against the position. As a result, they did not have to fight their confirmation bias and engage in the truly hard task of generating arguments against their own point of view. The resulting *myside bias* merely reflects a belief that it is better to provide arguments only for one’s side rather than also for the other side. As Wolfe and Britt (2008) observed, essayists often mention arguments for the other side – typically to rebut them and give even more weight to their own argument. But writing essays is as much of a basic argumentative skill as writing novels is a basic linguistic skill. It is therefore not surprising that untrained participants should exhibit a *myside bias* and that, as Wolfe points out, training can attenuate it.

The *belief bias* is one of the phenomena that, we surmised, show that people have a confirmation bias: They will take into account their beliefs about the conclusion when evaluating the logical validity of a syllogism. **De Neys and Stuppel & Ball** question our interpretation of the belief bias data. They both point out that people – at least some people – try to engage in logical reasoning when faced with such problems. That they try is hardly surprising: Participants are merely responding to instructions that emphasize logical validity. We agree that in reasoning tasks people try to provide the correct, logically valid answer. What is more interesting is that most of them fail. Given that the tasks are not computationally hard, this indicates that reasoning is not geared towards pure logical validity, but that it takes into account other factors, such as believability.

## R4. On the working of reasoning

### R4.1. The algorithmic level

Our target article focuses on the ultimate level of explanation: What is the function of reasoning? A theory at that level has implications for the algorithmic

implementation of reasoning, but this was not the subject matter of our article, and, to be candid, we are still in the process of working out these implications satisfactorily. At first blush, moreover, the argumentative approach does not imply a single, narrow characterization of this implementation. We therefore acknowledge this present limitation of our contribution that in particular **Khleutzos & Stevenson** have underscored, and we are grateful for the positive suggestions of several commentators.

**Oaksford's Bayesian probabilistic approach** (para. 4) allows for a fine-grained analysis of argument strength and of the changes in degree of beliefs they warrant. Like ours, this a computational theory, but it specifies a proximal function – how to change our beliefs in response to arguments – rather than an ultimate one. As a consequence, it has the potential of taking us one step further in the direction of an algorithmic theory.

**Dessalles** points out the sequential nature of reasoning, and asks how the argumentative theory can account for this feature. First, it is important to mention that even if the explicit part of reasoning is sequential – we make only one argument step at a time – other processes within reasoning (such as argument search) may well work in a parallel fashion. The sequential nature of explicit reasoning can be explained by its linguistic character (that might not be necessary for pure solitary reasoning if it exists, but that cannot be bypassed in argumentation), language itself being sequential, for several reasons unrelated to reasoning or argumentation (Pinker & Bloom 1990; see also Carruthers 1996).

**Wiles** mentions abduction as a plausible mechanism through which reasoning could find arguments. She points to other psychological mechanisms that also rely on abduction, such as spatial navigation, suggesting that reasoning might have been exapted from these mechanisms. However, given that reasoning deals with inputs and outputs that are very different from those of these other systems, the possibility of an evolutionary exaptation scenario remains very speculative. This does not mean that these other mechanisms have nothing to teach students of reasoning. To the extent that the problems solved by, for example, mechanisms of spatial navigation are similar to the problem facing reasoning – finding an acceptable argument within a large set of potentially relevant propositions – then, as suggested by Wiles's commentary, we can expect commonalities in the solutions used in both cases (for further elaboration on this point, see Mercier, submitted a).

**Uleman et al.** present interesting evidence and arguments on intuitive (or “spontaneous”) inferences. We agree that these inferences constitute most of cognition, and that they influence which arguments are used and how they are evaluated. We agree moreover that much more must be done on the relationship between intuitive inference and reasoning, even if much relevant work has already been done in the framework of dual-system approaches to reasoning.

Finally, **Weber & Johnson** offer a *process-level specification* of how reasoning works in decision making. According to their theory – *query theory* – participants facing a decision query their memory for relevant information or experiences. Two features of this theory are particularly relevant in the context of the argumentative theory. First, as a result of this process, “choice follows from the

resulting balance of evidence,” so that this theory predicts reason-based choice. The second important point is that “the first query produces richer representations because of output interference” (para. 5). To the extent that further queries might represent perspective that are less congenial to the individual – for instance, trying to see the problem from someone else's perspective – a mechanism that favors the first query can create a form of confirmation bias. Given that query theory predicts reason-based choice and that it might also explain some forms of confirmation bias, it is quite congenial to the argumentative approach. However, rather than being a question of “intrapsychic and interpersonal argumentation” (para. 8), as Weber & Johnson suggest, the question can perhaps be more usefully framed as a difference in level of analysis. We believe that query theory could also help explain aspects of interpersonal argumentation – how we find arguments – whereas the argumentative theory also makes predictions regarding intrapsychic phenomena – such as reason-based choice. The difference is that query theory offers predictions based on the *workings* of reasoning, whereas the argumentative theory offers predictions based on the *function* of reasoning. Given that they are not at the same level, these explanations do not compete. Assume, for instance, that reason-based choice is entirely due to the processes delineated in query theory. We would still need to understand why this process is there in the first place, and why the outcomes of such a process are generally adaptive – questions that the argumentative theory aims at answering.

Another way to better understand the processes of reasoning is through modeling. **Fox's Logic of Argument** is a possible solution that is congenial to our proposal because it distances itself from formal logic to encompass the subtleties of argument structure. The fields of AI and informal logic are now teeming with models of arguments that can be conveniently simulated on computers, and this is certainly an interesting way to develop for any theory of reasoning or argumentation.

#### **R4.2. Reasoning outside the lab**

The target article focused its review on experiments carried out in the laboratory, mostly with Western college students. **Narvaez** rightly points out the limitations of such a narrow focus. In their review of cross-cultural psychology work, Henrich et al. (2010) have shown that in many domains WEIRD people – people from western educated industrialized rich democratic countries – behave in ways that are different from the rest of the world (para. 9). In the case of reasoning and argumentation, scholars have hypothesized that such skills are a mostly Western tradition, born in classical Greece and nurtured in the Enlightenment. It would indeed be a deadly blow to the theory if some cultures were unwilling to argue or unable to reason. Happily for us (and for these cultures), the available data do not point in that direction. While there certainly are differences in reasoning and argumentative style (e.g., see Norenzayan et al. 2002), there is no report of a culture that would be deprived of these skills. The two most frequently alleged cases are illiterate societies – which are supposed to be unable to reason – and Eastern cultures – which are supposed to be unwilling to argue. Yet members

of illiterate populations can solve logical problems in the proper contexts, and the intellectual history of Eastern cultures is as fraught with debates as that of the West. While our theory could certainly be strengthened if the data from non-WEIRD cultures were richer, available data point to similar patterns the whole world over (Mercier, in press a).

**Narvaez** also regrets the absence of developmental data from the target article, accusing us of “implicit biological determinism” (para 5). Our thesis is evolutionary and pays attention to biological factors, but development itself is of obvious biological relevance. Again, language can be a useful analogy. Researchers who think that language is an adaptation do not deny that different languages are acquired in cognitive development. Likewise, the argumentative theory does not need argumentative skills to be just innate. The main reason so little space was devoted to an analysis of developmental evidence is that there is so much of it that it warranted a whole other article (Mercier, in press b). In this other article, it is shown that children can argue from very early on – long before they can do any abstract reasoning – that they are also prone to the confirmation bias, and that they reason better in groups – which is why collaborative learning has proven to be so successful in education.

**Narvaez** is joined by **Wu** in drawing attention to reasoning in the political sphere. The results discussed by Wu offer further support for the argumentative theory. In particular, he refers to the development of *deliberative democracy* as illustrating the power of group reasoning. Many arguments can be garnered in support of an increased participation of citizens in policy debates: more legitimate and fairer outcomes, increase in trust and positive public behavior, etc. But Wu points to a more recent trend that emphasizes the *epistemic* value of deliberative democracy: Beyond all their other advantages, debates can also yield epistemically superior outcomes (Cohen 1986; Estlund 2007; Landemore 2007). Empirical results in political science support the claim that groups can perform well, with deliberations yielding more informed and more coherent opinions and decisions (for review, see Mercier & Landemore, in press). Incidentally, the argumentative theory is also in a good position to help explain some of the blatant failures of political debates, whether it is polarization (Landemore & Mercier, submitted) or the dangers of widely publicized debates (Mercier, submitted b).

**Johnson** correctly argues that the theory should apply to scientists and philosophers – including the authors of this target article – as well as to laypersons. Indeed, it is not unlikely that, despite genuine efforts to remain objective, we have been guilty of the confirmation bias, thereby illustrating our argumentative theory by the very manner in which we were arguing for it. Argumentation and debates have always played a central role in philosophy, be it in classical Greece, Akbar’s India, or the Warring States Period in China. The lone philosopher always runs the risk of using her great reasoning abilities to build a system of high internal coherence on shaky intuitive foundations. Even if scientists rely more on empirical evidence than do philosophers to correct their intuitions, their reasoning is still deeply argumentative. A scientific article or book is always a “long argument” in Darwin’s famous words. Moreover – and contrary to popular representation of the lone genius –

groups have always been the crucial place for scientific reasoning (e.g., see Dunbar 1995).

**Johnson** also points out the importance of arguments in the moral domain, suggesting that people might be especially vulnerable to arguments that exploit moral “oughts.” Some moral psychologists would disagree and suggest instead that people are rarely receptive to moral arguments, being more easily influenced by narratives or emotional appeals (Bloom 2010; Haidt & Bjorklund 2007). The argumentative theory predicts an intermediary state of affairs: People should be somewhat receptive to moral arguments while evaluating them on the basis of their own moral intuitions (Mercier, in press c).

## R5. Conclusion

While our target article may have unwittingly offered an illustration of the confirmation bias, we hope that this discussion has, at least a little, exemplified the epistemic benefits of reasoning in group. It has not led us to revise the theory in any major way. Still, several commentaries point to fascinating directions for future research. More needs to be done to link our ultimate level theory with process theories of reasoning, and we are grateful for several very useful suggestions in this respect. We agree that reasoning outside the laboratory needs to be investigated more thoroughly and hope that a focus on argumentation and reasoning in interaction can help push in this direction. Finally, other mechanisms besides reasoning might benefit from being seen as having a social function. Ours is a contribution to the growing body of research showing how, and how much, the human mind is a social mind.

## References

[The letters “a” and “r” before author’s initials stand for target article and research references, respectively]

- Abelson, J., Forest, P.-G., Eyles, J., Smith, P., Martin, E. & Gauvin, F.-P. (2003) Deliberation about deliberative methods: Issues in the design and evaluation of public participation processes. *Social Science & Medicine* 57:239–51. [KC-CW]
- Acker, F. (2008) New findings on unconscious versus conscious thought in decision making: Additional empirical data and meta-analysis. *Judgment and Decision Making* 3(4):292–303. [aHM]
- Albrechtsen, J. S., Meissner, C. A. & Susa, K. J. (2009) Can intuition improve deception detection performance? *Journal of Experimental Social Psychology* 45(4):1052–55. [aHM]
- Allen, C., Bekoff, M. & Lauder, G., eds. (1998) *Nature’s purposes*. MIT Press. [aHM]
- Allport, F. (1924) *Social psychology*. Houghton Mifflin. [aHM]
- Ambady, N., Bernieri, F. J. & Richeson, J. A. (2000) Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In: *Advances in Experimental Social Psychology*, vol. 32, ed. M. P. Zanna, pp. 201–71. Academic Press. [aHM]
- Ambady, N. & Gray, H. (2002) On being sad and mistaken mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology* 83:947–61. [aHM]
- Ames, G. J. & Murray, F. B. (1982) When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology* 18:894–97. [JEO]
- Amir, O. & Ariely, D. (2003) Decision by rules: Disassociation between preferences and willingness to act. Working paper, Massachusetts Institute of Technology, Cambridge. [aHM]
- Anderson, C. A., Lepper, M. R. & Ross, L. (1980) Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology* 39(6):1037–49. [aHM]

- Anderson, C. A., New, B. L. & Speer, J. R. (1985) Argument availability as a mediator of social theory perseverance. *Social Cognition* 3(3):235–49. [aHM]
- Anderson, E. (2006) The epistemology of democracy. *Episteme: Journal of Social Epistemology* 3(1–2):8–22. [KC-CW]
- Anderson, T., Howe, C., Soden, R., Halliday, J. & Low, J. (2001) Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science* 29(1):1–32. [aHM]
- Anderson, T., Howe, C. & Tolmie, A. (1996) Interaction and mental models of physics phenomena: Evidence from dialogues between learners. In: *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird*, ed. J. Oakhill & A. Garnham, pp. 247–73. Psychology Press. [aHM]
- Ariely, D., Gneezy, U., Loewenstein, G. & Mazar, N. (2009) Large stakes and big mistakes. *Review of Economic Studies* 76(2):451–69. [aHM]
- Ariely, D. & Levav, J. (2000) Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of Consumer Research* 27(3):279–90. [aHM]
- Arkes, H. R. & Ayton, P. (1999) The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin* 125(5):591–600. [aHM]
- Arkes, H. R. & Blumer, C. (1985) The psychology of sunk cost. *Organizational Behavior and Human Decision Processes* 35(1):124–40. [aHM]
- Arkes, H. R., Guilmette, T. J., Faust, D. & Hart, K. (1988) Eliminating the hindsight bias. *Journal of Applied Psychology* 73(2):305–307. [aHM]
- Arleo, A. & Gerstner, W. (2000) Spatial cognition and neuro-mimetic navigation: A model of hippocampal place cell activity. *Biological Cybernetics* 83(3):287–99. [JW]
- Augustinova, M. (2008) Falsification cueing in collective reasoning: Example of the Wason selection task. *European Journal of Social Psychology* 38(5):770–85. [aHM]
- Bailenson, J. N. & Rips, L. J. (1996) Informal reasoning and burden of proof. *Applied Cognitive Psychology* 10(7):S3–16. [MH, arHM]
- Ball, L. J. (2010) The dynamics of reasoning: Chronometric analysis and dual-process theories. In: *The science of reason: A festschrift for Jonathan St. B. T. Evans*, ed. K. I. Manktelow, D. E. Over & S. Elqayam, pp. 283–307. Psychology Press. [EJNS]
- Ball, L. J., Phillips, P., Wade, C. N. & Quayle, J. D. (2006) Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology* 53:77–86. [WDN, EJNS]
- Bandura, A. (1990) Selective activation and disengagement of moral control. *Journal of Social Issues* 46(1):27–46. [aHM]
- Bandura, A., Barbaranelli, C., Caprara, G. V. & Pastorelli, C. (1996) Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology* 71:364–74. [aHM]
- Barber, B. M., Heath, C. & Odean, T. (2003) Good reasons sell: Reason-based choice among group and individual investors in the stock market. *Management Science* 49(12):1636–52. [aHM]
- Bar-Hillel, M. (1980) The base-rate fallacy in probability judgments. *Acta Psychologica* 44:211–33. [MO]
- Barkow, J. H., Cosmides, L. & Tooby, J., eds. (1992) *The adapted mind*. Oxford University Press. [aHM]
- Baron, J. (1994) Nonsequentialist decisions. *Behavioral and Brain Sciences* 17:1–42. [JStBTE]
- Baron, J. (1995) Myside bias in thinking about abortion. *Thinking & Reasoning* 1:221–35. [CRW]
- Barrouillet, P., Grosset, N. & Lecas, J.-F. (2000) Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition* 75:237–66. [JEO]
- Baumeister, R. F. (1997) *Evil: Inside human violence and cruelty*. Freeman. [aHM]
- Baumeister, R. F. (2005) *The cultural animal: Human nature, meaning, and social life*. Oxford University Press. [RFB]
- Baumeister, R. F. & Masicampo, E. J. (2010) Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychological Review* 117:945–71. [RFB]
- Baumeister, R. F., Masicampo, E. J. & Vohs, K. D. (2011) Do conscious thoughts cause behavior? *Annual Review of Psychology* 62:331–62. [RFB]
- Bazerman, M. H., Loewenstein, G. F. & White, S. B. (1992) Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly* 37(2):220–40. [aHM]
- Beatty, A. S., Reese, C. M., Persky, H. R. & Carr, P. (1996) NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress. U. S. Department of Education. Available at: <http://nces.ed.gov/pubsearch/pubinfo.asp?pubid=96085> [CRW]
- Bechara, A. (2005) Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience* 8:1458–63. [DN]
- Becker, G. & Stigler, G. J. (1977) De gustibus non est disputandum. *American Economic Review* 67:76–90. [EUW]
- Berger, J. A. & Heath, C. (2007) Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research* 34(2):121–34. [aHM]
- Bersoff, D. M. (1999) Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin* 25(1):28–39. [aHM]
- Besnard, P. & Hunter, A. (2008) *Elements of argumentation*. MIT Press. [JF]
- Billig, M. (1996) *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press. [DK, aHM]
- Blaisdell, A. P., Sawa, K., Leising, K. J. & Waldmann, M. R. (2006) Causal reasoning in rats. *Science* 311(5763):1020–22. [aHM]
- Blanchette, I. & Dunbar, K. (2001) Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition* 29(5):730–35. [aHM]
- Blinder, A. S. & Morgan, J. (2000) Are two heads better than one? An experimental analysis of group vs. individual decision making. NBER Working Paper 7909, National Bureau of Economic Research, Princeton, NJ. [aHM]
- Bloom, P. (2010) How do morals change? *Nature* 464(7288):490. [rHM]
- Blum-Kulka, S., Blondheim, M. & Hachohen, G. (2002) Traditions of dispute: From negotiations of Talmudic texts to the arena of political discourse in the media. *Journal of Pragmatics* 34(10–11):1569–94. [aHM]
- Boehm, C., with comments by Antweiler, C., Eibl-Eibesfeldt, I., Kent, S., Knauff, B. M., Mithen, S., Richerson, P. J. & Wilson, D. S. (1996) Emergency decisions, cultural-selection mechanics, and group selection. *Current Anthropology* 37(5):763–93. [aHM]
- Boiney, L. G., Kennedy, J. & Nye, P. (1997) Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes* 72(1):1–24. [aHM]
- Bond, S. D., Carlson, K. A., Meloy, M. G., Russo, J. E. & Tanner, R. J. (2007) Precommitment bias in the evaluation of a single option. *Organizational Behavior and Human Decision Processes* 102(2):240–54. [aHM]
- Bonner, B. L., Baumann, M. R. & Dalal, R. S. (2002) The effects of member expertise on group decision making and performance. *Organizational Behavior and Human Decision Processes* 88:719–36. [aHM]
- Bonner, C. & Newell, B. R. (2010) In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition* 38:186–96. [WDN]
- Bonner, S. E., Hastie, R., Sprinkle, G. B. & Young, S. M. (2000) A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research* 12(1):19–64. [aHM]
- Bonner, S. E. & Sprinkle, G. B. (2002) The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27(4–5):303–45. [aHM]
- Bragger, J. D., Hantula, D. A., Bragger, D., Kirman, J. & Kutcher, E. (2003) When success breeds failure: History, hysteresis, and delayed exit decisions. *Journal of Applied Psychology* 88(1):6–14. [aHM]
- Bragger, J. L., Bragger, D. H., Hantula, D. A. & Kirnan, J. P. (1998) Hysteresis and uncertainty: The effect of information on delays to exit decisions. *Organizational Behavior and Human Decision Processes* 74(3):229–53. [aHM]
- Braine, M. D. S. & O'Brien, D. P. (1998) *Mental logic*. Erlbaum. [JEO]
- Braman, E. (2009) *Law, politics, and perception: How policy preferences influence legal reasoning*. University of Virginia Press. [aHM]
- Bratman, M. E. (1987) *Intention, plans, and practical reason*. Harvard University Press. [KF]
- Brem, S. K. & Rips, L. J. (2000) Explanation and evidence in informal argument. *Cognitive Science* 24:573–604. [aHM]
- Briley, D. A., Morris, M. W. & Simonson, I. (2000) Reasons as carriers of culture: Dynamic versus dispositional models of cultural influence on decision making. *Journal of Consumer Research* 27(2):157–78. [aHM]
- Britt, M. A. & Kurby, C. A. (2005) Detecting incoherent informal arguments. Paper presented at the 15th Annual Meeting of the Society for Text and Discourse, Amsterdam, The Netherlands. [CRW]
- Britt, M. A., Kurby, C. & Wolfe, C. R. (2005) Memory for claims of simple arguments. Paper presented at the 15th Annual Meeting of the Society for Text and Discourse, Amsterdam, The Netherlands. [CRW]
- Britt, M. A., Kurby, C. A., Dandotkar, S. & Wolfe, C. R. (2008) I agreed with what? Memory for simple argument claims. *Discourse Processes* 45:52–84. [CRW]
- Britt, M. A. & Larson, A. A. (2003) Constructing representations of arguments. *Journal of Memory and Language* 48:794–810. [CRW]
- Brock, T. C. (1967) Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology* 3(3):269–309. [aHM]
- Brown, C. L. & Carpenter, G. S. (2000) Why is the trivial important? A reasons-based account for the effects of trivial attributes on choice. *Journal of Consumer Research* 26(4):372–85. [aHM]
- Brown, D. E. (1991) *Human universals*. McGraw-Hill. [aHM]
- Brownstein, A. L. (2003) Biased predecision processing. *Psychological Bulletin* 129(4):545–68. [aHM]

- Butera, F., Legrenzi, P., Mugny, G. & Pérez, J. A. (1992) Influence sociale et raisonnement. *Bulletin de Psychologie* 45:144–54. [aHM]
- Byrne, R. W. & Whiten, A., eds. (1988) *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford University Press. [aHM]
- Cacioppo, J. T. & Petty, R. E. (1979) Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology* 37(1):97–109. [aHM]
- Camerer, C. & Hogarth, R. M. (1999) The effect of financial incentives on performance in experiments: A review and capital-labor theory. *Journal of Risk and Uncertainty* 19(1):7–42. [aHM]
- Campion, N. (2006) Hypothetical and certain inferences from conditional arguments read in texts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:547–58. [JSU]
- Carlston, D. E. & Skowronski, J. J. (2005) Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology* 89:884–98. [JSU]
- Carpenter, G. S., Glazer, R. & Nakamoto, K. (1994) Meaningful brand from meaningless differentiation: The dependence on irrelevant attributes. *Journal of Marketing Research* 31(3):339–50. [aHM]
- Carruthers, P. (1996) *Language, thought and consciousness: An essay in philosophical psychology*. Cambridge University Press. [KF, rHM]
- Carruthers, P. (1998) Conscious thinking: Language or elimination? *Mind and Language* 13:457–76. [KF]
- Chaiken, S. & Yates, S. (1985) Affective-cognitive consistency and thought-induced attitude polarization. *Journal of Personality and Social Psychology* 49(6):1470–81. [aHM]
- Chater, N. & Oaksford, M. (1999) The probability heuristics model of syllogistic reasoning. *Cognitive Psychology* 38:191–258. [aHM]
- Chernev, A. (2005) Context effects without a context: Attribute balance as a reason for choice. *Journal of Consumer Research* 32(2):213–23. [aHM]
- Christensen-Szalanski, J. J. & Beach, L. R. (1984) The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist* 39(1):75–78. [aHM]
- Claxton, G. (1997) *Hare brain, tortoise mind: How intelligence increases when you think less*. HarperCollins. [aHM]
- Clément, F. (2010) To trust or not to trust? Children's social epistemology. *Review of Philosophy and Psychology* 1(4):531–49. [aHM]
- Cohen, J. (1986) An epistemic conception of democracy. *Ethics* 97(1):26–38. [rHM]
- Connolly, T., Ordóñez, L. D. & Coughlan, R. (1997) Regret and responsibility in the evaluation of decision outcomes. *Organizational Behavior and Human Decision Processes* 70:73–85. [TC]
- Connolly, T., Reb, J. & Kausel, E. E. (2010) Intuitive politicians or intuitive penitents? Regret aversion, accountability and justification in the decoy effect. Working paper, University of Arizona, Tucson. [TC]
- Connolly, T. & Zeelenberg, M. (2002) Regret in decision making. *Current Directions in Psychological Science* 11:212–16. [TC]
- Corner, A. & Hahn, U. (2009) Evaluating science arguments: Evidence, uncertainty, and argument strength. *Journal of Experimental Psychology: Applied* 15(3):199–212. [aHM]
- Corner, A., Hahn, U. & Oaksford, M. (2006) The slippery slope argument: Probability, utility and category reappraisal. In: *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*, ed. R. Sun & N. Miyake, pp. 1145–50. Erlbaum. [aHM]
- Cowley, M. & Byrne, R. M. J. (2005) When falsification is the only path to truth. In: *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, ed. B. G. Bara, L. Barsalou & M. Buchiarielli, pp. 512–17. Erlbaum. [aHM, FHP]
- Crain, S. & Khlentzos, D. (2010) The logic instinct. *Mind and Language* 25(1):30–65. [DMK]
- Crandall, C. S. & Eshleman, A. (2003) A justification–suppression model of the expression and experience of prejudice. *Psychological Bulletin* 129(3):414–46. [aHM]
- Crosan, R. T. A. (1999) The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes* 80(2):118–33. [aHM]
- Csikszentmihalyi, M. & Sawyer, R. K. (1995) Creative insight: The social dimension of a solitary moment. In: *The nature of insight*, ed. R. J. Sternberg & J. E. Davidson, pp. 329–63. MIT Press. [aHM]
- Cunningham, C. B., Schilling, N., Anders, C. & Carrier, D. R. (2010) The influence of foot posture on the cost of transport in humans. *Journal of Experimental Biology* 213(5):790–97. [aHM]
- Dana, J., Weber, R. A. & Kuang, J. X. (2007) Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1):67–80. [aHM]
- Das, S., Fox, J., Elsdon, D. & Hammond, P. (1997) A flexible architecture for a general intelligent agent. *Journal of Experimental & Theoretical Artificial Intelligence* 9:407–40. [JF]
- Davies, M. F. (1992) Field dependence and hindsight bias: Cognitive restructuring and the generation of reasons. *Journal of Research in Personality* 26(1):58–74. [aHM]
- Davis, J. H. (1973) Group decisions and social interactions: A theory of social decision schemes. *Psychological Review* 80(2):97–125. [aHM]
- Dawkins, R. & Krebs, J. R. (1978) Animal signals: Information or manipulation? In: *Behavioural ecology: An evolutionary approach*, ed. J. R. Krebs & N. B. Davies, pp. 282–309. Basil Blackwell. [aHM]
- Dawson, E., Gilovich, T. & Regan, D. T. (2002) Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin* 28(10):1379–87. [aHM]
- De Neys, W. & Franssens, S. (2009) Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* 113:45–61. [WDN]
- De Neys, W. & Glumicic, T. (2008) Conflict monitoring in dual process theories of reasoning. *Cognition* 106:1248–99. [WDN]
- De Neys, W., Moyens, E. & Vansteenwegen, D. (2010) Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, & Behavioral Neuroscience* 10:208–16. [WDN]
- De Neys, W., Vartanian, O. & Goel, V. (2008) Smarter than we think: When our brains detect that we are biased. *Psychological Science* 19:483–89. [WDN]
- Dennett, D. C. (1969) *Content and consciousness*. Routledge & Kegan Paul. [aHM]
- Dessalles, J.-L. (1998) Altruism, status, and the origin of relevance. In: *Approaches to the evolution of language: Social and cognitive bases*, ed. J. R. Hurford, M. Studdert-Kennedy & C. Knight, pp. 130–47. Cambridge University Press. Available at: [http://www.dessalles.fr/papiers/pap.evol/Dessalles\\_96122602.pdf](http://www.dessalles.fr/papiers/pap.evol/Dessalles_96122602.pdf) [J-LD]
- Dessalles, J.-L. (2007) *Why we talk: The evolutionary origins of language*. Oxford University Press. [J-LD, aHM]
- Dessalles, J.-L. (2008) *La pertinence et ses origines cognitives: Nouvelles théories*. Hermes Science. Available at: <http://pertinence.dessalles.fr> [J-LD]
- DeWall, C. N., Baumeister, R. F. & Masicampo, E. J. (2008) Evidence that logical reasoning depends on conscious processing. *Consciousness and Cognition* 17:628–45. [RFB]
- Diekmann, K. A., Samuels, S. M., Ross, L. & Bazerman, M. H. (1997) Self-interest and fairness in problems of resource allocation: Allocators versus recipients. *Journal of Personality and Social Psychology* 72(5):1061–74. [aHM]
- Dijksterhuis, A. (2004) Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology* 87(5):586–98. [aHM]
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F. & van Baaren, R. B. (2006a) Complex choices better made unconsciously? *Science* 313:760–61. [EUW]
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F. & van Baaren, R. B. (2006b) On making the right choice: The deliberation-without-attention effect. *Science* 311(5763):1005–1007. [aHM, JStBTE]
- Dijksterhuis, A., Bos, M. W., van der Leij, A. & van Baaren, R. B. (2009) Predicting soccer matches after unconscious and conscious thought as a function of expertise. *Psychological Science* 20(11):1381–87. [aHM]
- Dijksterhuis, A. & van Olden, Z. (2006) On the benefits of thinking unconsciously: Unconscious thought can increase post-choice satisfaction. *Journal of Experimental Social Psychology* 42(5):627–31. [aHM]
- Ditto, P. H. & Lopez, D. F. (1992) Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology* 63(4):568–84. [aHM]
- Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepanky, J. A. & Lockhart, L. K. (2003) Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin* 29(9):1120–32. [aHM]
- Ditto, P. H., Scepanky, J. A., Munro, G. D., Apanovitch, A. M. & Lockhart, L. K. (1998) Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology* 75(1):53–69. [aHM]
- Dubreuil, B. (2010) Paleolithic public goods games: Why human culture and cooperation did not evolve in one step. *Biology and Philosophy* 25(1):53–73. [aHM]
- Dunbar, K. (1995) How scientists really reason: Scientific reasoning in real-world laboratories. In: *The nature of insight*, ed. R. J. Steinberg & J. Davidson, pp. 365–95. MIT Press. [rHM]
- Dunbar, K. (1997) How scientists think: Online creativity and conceptual change in science. In: *Conceptual structures and processes: Emergence discovery and change*, ed. T. B. Ward, S. M. Smith & S. Vaid, pp. 461–93. American Psychological Association. [aHM]
- Dunbar, R. I. M. (1996) The social brain hypothesis. *Evolutionary Anthropology* 6:178–90. [aHM]
- Dunbar, R. I. M. & Shultz, S. (2003) Evolution of the social brain. *Science* 302:1160–61. [aHM]
- Dung, P. M. (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. *Artificial*



- Intelligence* 77:321–57. Available at: <http://dli.iit.ac.in/tjcai/IJCAI-93-VOL2/PDF/003.pdf> [J-LD]
- Dunning, D., Meyerowitz, J. A. & Holzberg, A. D. (1989) Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology* 57(6):1082–90. [aHM]
- Eagly, A. H., Kulesa, P., Brannon, L. A., Shaw, K. & Hutson-Comeaux, S. (2000) Why counterattitudinal messages are as memorable as proattitudinal messages: The importance of active defense against attack. *Personality and Social Psychology Bulletin* 26(11):1392–408. [aHM]
- Ebbesen, E. B. & Bowers, R. J. (1974) Proportion of risky to conservative arguments in a group discussion and choice shifts. *Journal of Personality and Social Psychology* 29(3):316–27. [aHM]
- Edwards, K. & Smith, E. E. (1996) A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology* 71(1):5–24. [aHM]
- Elvang-Goransson, M., Krause, P. J. & Fox, J. (1993) Acceptability of arguments as logical uncertainty. In: *Symbolic and quantitative approaches to reasoning and uncertainty: Proceedings, European Conference ECSQUARU 93*, ed. M. Clarke, R. Kruse & S. Moral, pp. 79–84. Springer-Verlag. [JF]
- Esser, J. K. (1998) Alive and well after 25 years: A review of groupthink research. *Organizational Behavior and Human Decision Processes* 73(2–3):116–41. [aHM]
- Esser, J. K. & Lindoerfer, J. S. (1989) Groupthink and the space shuttle Challenger accident: Toward a quantitative case analysis. *Journal of Behavioral Decision Making* 2(3):167–77. [aHM]
- Estlund, D. (2007) *Democratic authority. A philosophical framework*. Princeton University Press. [rHM]
- Evans, J. St. B. T. (1989) *Bias in human reasoning: Causes and consequences*. Erlbaum. [aHM]
- Evans, J. St. B. T. (1996) Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology* 87:223–40. [aHM]
- Evans, J. St. B. T. (2000) Thinking and believing. In: *Mental models in reasoning*, ed. J. García-Madruga, N. Carriedo & M. J. González-Labra, pp. 41–56. Universidad Nacional de Educación a Distancia. [EJNS]
- Evans, J. St. B. T. (2002) Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin* 128(6):978–96. [aHM]
- Evans, J. St. B. T. (2007) *Hypothetical thinking: Dual processes in reasoning and judgment*. Psychology Press. [JStBTE, aHM, EJNS]
- Evans, J. St. B. T. (2008) Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology* 59:255–78. [JStBTE]
- Evans, J. St. B. T. (2009) How many dual process theories do we need: One, two or many? In: *In two minds: Dual processes and beyond*, ed. J. St. B. T. Evans & K. Frankish, pp. 33–54. Oxford University Press. [EJNS]
- Evans, J. St. B. T. (2010) *Thinking twice: Two minds in one brain*. Oxford University Press. [JStBTE]
- Evans, J. St. B. T. & Ball, L. J. (2010) Do people reason on the Wason selection task? A new look at the data of Ball et al. (2003). *Quarterly Journal of Experimental Psychology* 63:434–41. [JStBTE]
- Evans, J. St. B. T., Barston, J. L. & Pollard, P. (1983) On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition* 11:295–306. [aHM]
- Evans, J. St. B. T., Handley, S. J., Harper, C. N. J. & Johnson-Laird, P. N. (1999) Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(6):1495–513. [aHM]
- Evans, J. St. B. T. & Lynch, J. S. (1973) Matching bias in the selection task. *British Journal of Psychology* 64(3):391–97. [aHM]
- Evans, J. St. B. T., Newstead, S. E. & Byrne, R. M. J. (1993) *Human reasoning: The psychology of deduction*. Erlbaum. [aHM, JEO]
- Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning*. Psychology Press. [JStBTE, aHM]
- Evans, J. St. B. T. & Wason, P. C. (1976) Rationalisation in a reasoning task. *British Journal of Psychology* 63:205–12. [JStBTE, aHM]
- Farnsworth, P. R. & Behner, A. (1931) A note on the attitude of social conformity. *Journal of Social Psychology* 2:126–28. [aHM]
- Fearon, J. D. (1998) Deliberation as discussion. In: *Deliberative democracy*, ed. J. Elster, pp. 44–68. Cambridge University Press. [KC-CW]
- Fiedler, K., Schenck, W., Watling, M. & Menges, J. I. (2005) Priming trait inferences through pictures and moving pictures: The impact of open and closed mindsets. *Journal of Personality and Social Psychology* 88:229–44. [JSU]
- Foot, H., Howe, C., Anderson, A., Tormie, A. & Warden, D. (1994) *Group and interactive learning*. Computational Mechanics Press. [aHM]
- Fox, J. (1980) Making decisions under the influence of memory. *Psychological Review* 87(2):190–211. [JF]
- Fox, J. (2003) Logic, probability and the cognitive foundations of rational belief. *Journal of Applied Logic* 1:197–224. [JF]
- Fox, J. (in press) Arguing about the evidence. In: *Evidence, inference and enquiry*, ed. P. Dawid, W. Twining & M. Vasilaki. Oxford University Press/British Academy. [JF]
- Fox, J., Beveridge, M. & Glasspool, D. (2003) Understanding intelligent agents: Analysis and synthesis. *AI Communications* 16(3):139–52. [JF]
- Fox, J., Clark, D., Glowinski, A. & O’Neil, M. (1990) Using predicate logic to integrate qualitative reasoning and classical decision theory. *IEEE Transactions on Systems, Man, and Cybernetics* 20(2):347–57. [JF]
- Fox, J. & Das, S. (2000) *Safe and sound: Artificial intelligence in hazardous applications*. MIT Press. [JF]
- Fox, J., Glasspool, D., Grecu, D., Modgil, S., South, M. & Patkar, V. (2007) Argumentation-based inference and decision-making: A medical perspective. *IEEE Intelligent Systems* 22(6):34–41. [JF]
- Fox, J., Krause, P. J. & Elvang-Goransson, M. (1993) Argumentation as a general framework for uncertain reasoning. In: *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence*, ed. D. Heckerman and E. H. Mamdani, pp. 428–34. Morgan Kaufman. [JF]
- Frankish, K. (1998) Natural language and virtual belief. In: *Language and thought: Interdisciplinary themes*, ed. P. Carruthers & J. Boucher, pp. 248–69. Cambridge University Press. [KF]
- Frankish, K. (2004) *Mind and supermind*. Cambridge University Press. [KF]
- Frankish, K. (2009) Systems and levels: Dual-system theories and the personal-subpersonal distinction. In: *In two minds: Dual processes and beyond*, ed. J. St. B. T. Evans & K. Frankish, pp. 89–107. Oxford University Press. [KF]
- Franklin, B. (1817/2006) *The autobiography of Benjamin Franklin*. NuVision. (Original work published 1817.) [aHM, EUW]
- Frederick, S. (2005) Cognitive reflection and decision making. *Journal of Economic Perspectives* 19(4):25–42. [JStBTE]
- Fricker, M. (2007) *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press. [KC-CW]
- Fry, D. P. (2006) *The human potential for peace: An anthropological challenge to assumptions about war and violence*. Oxford University Press. [DN]
- Garland, H. (1990) Throwing good money after bad: The effect of sunk costs on the decision to escalate commitment to an ongoing project. *Journal of Applied Psychology* 75(6):728–31. [aHM]
- Geurts, B. (2003) Reasoning with quantifiers. *Cognition* 86(3):223–51. [aHM]
- Gibbard, A. (1990) *Wise choices, apt feelings*. Cambridge University Press. [aHM]
- Gigerenzer, G. (2007) *Gut feelings*. Penguin. [JStBTE]
- Gilbert, D. T. (2002) Inferential correction. In: *Heuristics and biases*, ed. T. Gilovich, D. Griffin & D. Kahneman, pp. 167–84. Cambridge University Press. [aHM]
- Gilbert, D. T. & Ebert, J. E. J. (2002) Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology* 82(4):503–14. [aHM]
- Gilovich, T. (1983) Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology* 44(6):1110–26. [aHM]
- Giroto, V., Kimmelmeier, M., Sperber, D. & Van der Henst, J.-B. (2001) Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition* 81(2):69–76. [aHM]
- Gladwell, M. (2005) *Blink: The power of thinking without thinking*. Little, Brown. [aHM]
- Gladwell, M. (2005) *Blink*. Penguin. [JStBTE]
- Godden, D. M. & Walton, D. (2004) Denying the antecedent as a legitimate argumentative strategy: A dialectical model. *Informal Logic* 24:219–43. [MO]
- Godlee, F. (2010) Conflicts of interest and pandemic flu: WHO must act now to restore its credibility, and Europe should legislate. *British Medical Journal* 340:1256–57. [KC-CW]
- Goldberg, E. (2001) *The executive brain*. Oxford University Press. [DN]
- Goldstein, M., Crowell, A. & Kuhn, D. (2009) What constitutes skilled argumentation and how does it develop? *Informal Logic* 29(4):379–95. [DK, rHM]
- Goren, A. & Todorov, A. (2009) Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition* 27:222–48. [JSU]
- Graff, G. (2003) *Clueless in academe: How schooling obscures the life of the mind*. Yale University Press. [DK]
- Green, K. C., Armstrong, J. C. & Graefe, A. (2007) Methods to elicit forecasts from groups: Delphi and prediction markets compared. *Foresight: The International Journal of Applied Forecasting* Fall: 17–21. [aHM]
- Greene, J. D. (2003) From neural “is” to moral “ought”: What are the moral implications of neuroscientific moral psychology? *Nature Reviews: Neuroscience* 4:847–50. [JA]
- Greenwald, A. G. (1969) The open-mindedness of the counterattitudinal role player. *Journal of Experimental Social Psychology* 5(4):375–88. [aHM]
- Greenwald, E. A., Persky, H. R., Campbell, J. R. & Mazzeo, J. National Assessment of Educational Progress. (1999) *NAEP 1998 Writing Report Card for the Nation and the States*. U. S. Department of Education. Available at: <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=19999462>. [CRW]
- Grice, H. P. (1975) Logic and conversation. In: *Syntax and semantics, vol. 3: Speech acts*, ed. P. Cole & J. P. Morgan. Seminar Press. [aHM]
- Griffin, D. W. & Dunning, D. (1990) The role of construal processes in overconfident predictions about the self and others. *Journal of Personality* 59(6):1128–39. [aHM]

- Guenther, C. L. & Alicke, M. D. (2008) Self-enhancement and belief perseverance. *Journal of Experimental Social Psychology* 44(3):706–12. [aHM]
- Gummerum, M., Keller, M., Takezawa, M. & Mata, J. (2008) To give or not to give: Children's and adolescents' sharing and moral negotiations in economic decision situations. *Child Development* 79(3):562–76. [aHM]
- Gutmann, A. & Thompson, D. (2004) *Why deliberative democracy?* Princeton University Press. [KC-CW]
- Hafer, C. L. & Begue, L. (2005) Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin* 131(1):128–67. [aHM]
- Hagler, D. A. & Brem, S. K. (2008) Reaching agreement: The structure & pragmatics of critical care nurses' informal argument. *Contemporary Educational Psychology* 33(3):403–24. [aHM]
- Hahn, U. & Oaksford, M. (2007) The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review* 114:704–32. [aHM, MO]
- Hahn, U., Oaksford, M. & Bayindir, H. (2005) How convinced should we be by negative evidence? In: *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, ed. B. G. Bara, L. Barsalou & M. Buchiarielli, pp. 887–92. Erlbaum. [aHM]
- Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4):814–34. [JA, aHM]
- Haidt, J. & Bjorklund, F. (2007) Social intuitionists reason, in conversation. In: *Moral Psychology, vol. 2: The cognitive science of morality: Intuition and diversity*, ed. W. Sinnott-Armstrong, pp. 241–54. MIT Press. [arHM]
- Halberstadt, J. B. & Levine, G. M. (1999) Effects of reasons analysis on the accuracy of predicting basketball games. *Journal of Applied Social Psychology* 29(3):517–30. [aHM]
- Ham, J. & van den Bos, K. (2008) Not fair for me! The influence of personal relevance on social justice inferences. *Journal of Experimental Social Psychology* 44:699–705. [JSU]
- Ham, J. & Vonk, R. (2003) Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology* 39:434–47. [JSU]
- Hamilton, R. W. & Thompson, D. V. (2007) Is there a substitute for direct experience? Comparing consumers' preferences after direct and indirect product experiences. *Journal of Consumer Research* 34(4):546–55. [aHM]
- Hardisty, D. H., Johnson, E. J. & Weber, E. U. (2010) A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. *Psychological Science* 21:86–92. [EUW]
- Harman, G. (1986) *Change in view: Principles of reasoning*. MIT Press. [aHM]
- Harrell, M. (2006) Diagrams that really are worth ten thousand words: Using argument diagrams to teach critical thinking skills. In: *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, p. 2501. Erlbaum. [MH]
- Harrell, M. (2008) No computer program required: Even pencil-and-paper argument mapping improves critical thinking skills. *Teaching Philosophy* 31:351–74. [MH]
- Harrell, M. (forthcoming) Argument diagramming and critical thinking in introductory philosophy. *Higher Education Research and Development*. [MH]
- Harris, P. L. (2007) Trust. *Developmental Science* 10(1):135–38. [aHM]
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, L., Lindberg, M. & Merrill, L. (2009) Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin* 135(4):555–88. [aHM]
- Hassin, R. R., Bargh, J. A. & Uleman, J. S. (2002) Spontaneous causal inferences. *Journal of Experimental Social Psychology* 38:515–22. [JSU]
- Henrich, J., Heine, S. & Norenzayan, A. (2010) The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3):61–83. [rHM, DN]
- Hill, G. W. (1982) Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin* 91(3):517–39. [aHM]
- Hinsz, V. B., Tindale, R. S. & Nagao, D. H. (2008) Accentuation of information processes and biases in group judgments integrating base-rate and case-specific information. *Journal of Experimental Social Psychology* 44(1):116–26. [aHM]
- Hirt, E. R. & Markman, K. D. (1995) Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology* 69(6):1069–86. [aHM]
- Hoch, S. J. (1985) Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11(4):719–31. [aHM]
- Hogarth, R. M. (2001) *Educating intuition*. University of Chicago Press. [DN]
- Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B. & Tzourio-Mazoyer, N. (2000) Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience* 12:721–28. [WDN]
- Howe, C. J. (1990) Physics in the primary school: Peer interaction and the understanding of floating and sinking. *European Journal of Psychology of Education* 5(4):459–75. [aHM]
- Hrdy, S. B. (2009) *Mothers and others*. Belknap Press. [aHM]
- Hsee, C. K. (1995) Elastic justification: How tempting but task-irrelevant factors influence decisions. *Organizational Behavior and Human Decision Processes* 62(3):330–37. [aHM]
- Hsee, C. K. (1996a) Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes* 66(1):122–29. [aHM]
- Hsee, C. K. (1996b) The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes* 67(3):247–57. [aHM]
- Hsee, C. K. (1998) Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making* 11(2):107–21. [aHM]
- Hsee, C. K. (1999) Value seeking and prediction-decision inconsistency: Why don't people take what they predict they'll like the most? *Psychonomic Bulletin & Review* 6(4):555–61. [aHM]
- Hsee, C. K. & Hastie, R. (2006) Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences* 10(1):31–37. [aHM]
- Hsee, C. K., Loewenstein, G. F., Blount, S. & Bazerman, M. H. (1999) Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin* 125(5):576–90. [aHM]
- Hsee, C. K. & Zhang, J. (2004) Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology* 86(5):680–95. [aHM]
- Hsee, C. K., Zhang, J., Yu, F. & Xi, Y. (2003) Lay rationalism and inconsistency between predicted experience and decision. *Journal of Behavioral Decision Making* 16(4):257–72. [aHM]
- Huber, J., Payne, J. W. & Puto, C. (1982) Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research* 9(1):90–98. [aHM]
- Humphrey, N. K. (1976) The social function of Intellect. In: *Growing points in ethology*, ed. P. P. G. Bateson & R. A. Hinde, pp. 303–17. Cambridge University Press. [aHM]
- Igou, E. R. (2004) Lay theories in affective forecasting: The progression of affect. *Journal of Experimental Social Psychology* 40(4):528–34. [aHM]
- Igou, E. R. & Bless, H. (2007) On undesirable consequences of thinking: Framing effects as a function of substantive processing. *Journal of Behavioral Decision Making* 20(2):125–42. [aHM]
- Irwin, J. R., Slovic, P., Lichtenstein, S. & McClelland, G. H. (1993) Preference reversals and the measurement of environmental values. *Journal of Risk and Uncertainty* 6(1):5–18. [aHM]
- Isenberg, D. J. (1986) Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology* 50(6):1141–51. [aHM]
- Jackendoff, R. (1996) How language helps us think. *Pragmatics and Cognition* 4(1):1–34. [aHM]
- Janis, I. L. (1972) *Victims of groupthink*. Houghton-Mifflin. [RJS]
- Janis, I. L. (1982) *Groupthink*, 2nd rev. ed. Houghton Mifflin. [aHM]
- Janis, I. L. & Mann, L. (1977) *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press. [aHM]
- Jananoff, S. (2003) Technologies of humility: Citizen participation in governing science. *Minerva* 41:223–44. [KC-CW]
- Jaswal, V. K. & Neely, L. A. (2006) Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science* 17:757–58. [JEO]
- Jellison, J. M. & Mills, J. (1969) Effect of public commitment upon opinions. *Journal of Experimental Social Psychology* 5(3):340–46. [aHM]
- Johnson, D. W. & Johnson, R. T. (2007) *Creative constructive controversy: Intellectual challenge in the classroom*, 4th ed. Interaction. [aHM]
- Johnson, D. W. & Johnson, R. T. (2009) Energizing learning: The instructional power of conflict. *Educational Researcher* 38(1):37–51. [aHM]
- Johnson, E. J. & Goldstein, D. G. (2003) Do defaults save lives? *Science* 302:1338–39. [EUW]
- Johnson, E. J., Haubl, G. & Keinan, A. (2007) Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33:461–73. [aHM, EUW]
- Johnson, J. A. (2007, June) *The evolution of moral rules from natural laws*. Poster presented at the 19th annual meeting of the Human Behavior and Evolution Society, Williamsburg, VA. [JA]
- Johnson, J. A., Germer, C. K., Efran, J. S. & Overton, W. F. (1988) Personality as the basis for theoretical predilections. *Journal of Personality and Social Psychology* 55:824–35. [JA]
- Johnson-Laird, P. N. (2006) *How we reason*. Oxford University Press. [arHM]
- Johnson-Laird, P. N. & Byrne, R. M. J. (2002) Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review* 109(4):646–78. [aHM]
- Johnson-Laird, P. N. & Wason, P. C. (1970) Insight into a logical relation. *Quarterly Journal of Experimental Psychology* 22(1):49–61. [aHM]
- John-Steiner, V. (2000) *Creative collaboration*. Oxford University Press. [aHM]

- Jones, M. & Sugden, R. (2001) Positive confirmation bias in the acquisition of information. *Theory and Decision* 50(1):59–99. [aHM]
- Kahneman, D. (2003) A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58(9):697–720. [aHM]
- Kahneman, D. & Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgement. In: *Heuristics and biases: The psychology of intuitive judgment*, ed. T. Gilovich, D. Griffin & D. Kahneman, pp. 49–81. Cambridge University Press. [JStBTE, aHM]
- Kahneman, D. & Frederick, S. (2005) A model of heuristic judgment. In: *The Cambridge handbook of thinking and reasoning*, ed. K. Holyoak & R. G. Morrison, pp. 267–94. Cambridge University Press. [aHM]
- Kahneman, D. & Ritov, I. (1994) Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty* 9(1):5–37. [aHM]
- Kahneman, D. & Tversky, A. (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3):430–54. [aHM]
- Kahneman, D. & Tversky, A. (1973) On the psychology of prediction. *Psychological Review* 80:237–57. [MO]
- Kahneman, D., Slovic, P. & Tversky, A. (1982) *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. [aHM]
- Kaplan, M. F. & Miller, C. E. (1977) Judgments and group discussion: Effect of presentation and memory factors on polarization. *Sociometry* 40(4):337–43. [aHM]
- Katz, J. J. (1986) *Cogitations*. Oxford University Press. [aHM]
- Keeney, S., Hasson, F. & McKenna, H. P. (2001) A critical review of the Delphi technique as a research methodology for nursing. *International Journal of Nursing Studies* 38(2):195–200. [aHM]
- Kerr, N. L., Maccoun, R. J. & Kramer, G. P. (1996) Bias in judgment: Comparing individuals and groups. *Psychological Review* 103(4):687–719. [aHM]
- Kerr, N. L. & Tindale, R. S. (2004) Group performance and decision making. *Annual Review of Psychology* 55:623–55. [aHM]
- Kersten, D., Mamassian, P. & Yuille, A. (2004) Object perception as Bayesian inference. *Annual Review of Psychology* 55:271–304. [aHM]
- Klaczynski, P. A. (1997) Bias in adolescents' everyday reasoning and its relationship with intellectual ability, personal theories, and self-serving motivation. *Developmental Psychology* 33(2):273–83. [aHM]
- Klaczynski, P. A. & Cottrell, J. M. (2004) A dual-process approach to cognitive development: The case of children's understanding of sunk cost decisions. *Thinking & Reasoning* 10(2):147–74. [aHM]
- Klaczynski, P. A. & Gordon, D. H. (1996a) Everyday statistical reasoning during adolescence and young adulthood: Motivational, general ability, and developmental influences. *Child Development* 67(6):2873–91. [aHM]
- Klaczynski, P. A. & Gordon, D. H. (1996b) Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology* 62(3):317–39. [aHM]
- Klaczynski, P. A., Gordon, D. H. & Fauth, J. (1997) Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology* 89(3):470–85. [aHM]
- Klaczynski, P. A. & Lavalley, K. L. (2005) Domain-specific identity, epistemic regulation, and intellectual ability as predictors of belief-based reasoning: A dual-process perspective. *Journal of Experimental Child Psychology* 92(1):1–24. [aHM]
- Klaczynski, P. A. & Narasimham, G. (1998) Development of scientific reasoning biases: Cognitive versus ego-protective explanations. *Developmental Psychology* 34(1):175–87. [aHM]
- Klaczynski, P. A. & Robinson, B. (2000) Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning tasks. *Psychology and Aging* 15(3):400–16. [aHM]
- Klauer, K. C., Musch, J. & Naumer, B. (2000) On belief bias in syllogistic reasoning. *Psychological Review* 107(4):852–84. [aHM, EJNS]
- Klayman, J. & Ha, Y.-W. (1987) Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review* 94(2):211–28. [aHM, FHP]
- Klein, G. (1998) *Sources of power: How people make decisions*. MIT Press. [aHM]
- Koehler, J. J. (1993) The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes* 56(1):28–55. [aHM]
- Koenig, M. A., Clement, F. & Harris, P. L. (2004) Trust in testimony: Children's use of true and false statements. *Psychological Science* 15:694–98. [JEO]
- Kogan, N. & Wallach, M. A. (1966) Modification of a judgmental style through group interaction. *Journal of Personality and Social Psychology* 4(2):165–74. [aHM]
- Konrath, S., O'Brien, E. H. & Hsing, C. (in press) Changes in dispositional empathy over time in college students: A meta-analysis. *Personality and Social Psychology Review*. [DN]
- Koole, S. L., Dijksterhuis, A. & Van Knippenberg, A. (2001) What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology* 80(4):669–85. [aHM]
- Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980) Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6(2):107–18. [aHM]
- Kray, L. & Gonzalez, R. (1999) Differential weighting in choice versus advice: I'll do this, you do that. *Journal of Behavioral Decision Making* 12(3):207–17. [aHM]
- Krebs, J. R. & Dawkins, R. (1984) Animal signals: Mind-reading and manipulation? In: *Behavioural ecology: An evolutionary approach*, 2nd ed., ed. J. R. Krebs & N. B. Davies, pp. 390–402. Basil Blackwell. [aHM]
- Kressel, L. (2010) *Spontaneous trait inferences, and explicit and implicit theories about traits causing behaviors*. Unpublished manuscript, New York University. [JSU]
- Kressel, L. & Uleman, J. S. (2010) Personality traits function as causal concepts. *Journal of Experimental Social Psychology* 46:213–16. [JSU]
- Kruglanski, A. W. & Freund, T. (1983) The freezing and unfreezing of lay-inferences: Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology* 19(5):448–68. [aHM]
- Kuhn, D. (1991) *The skills of argument*. Cambridge University Press. [DK, aHM]
- Kuhn, D. (1992) Thinking as argument. *Harvard Educational Review* 62(2):155–78. [aHM]
- Kuhn, D. (2001) How do people know? *Psychological Sciences* 12:1–8. [CRW]
- Kuhn, D. & Crowell, A. (in press) Argumentation as a path to the thinking development of young adolescents. *Psychological Science*. [DK]
- Kuhn, D., Goh, W., Iordanou, K. & Shaenfield, D. (2008) Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development* 79(5):1310–29. [DK, rHM]
- Kuhn, D. & Lao, J. (1996) Effects of evidence on attitudes: Is polarization the norm? *Psychological Science* 7(2):115–20. [aHM]
- Kuhn, D., Shaw, V. F. & Felton, M. (1997) Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction* 15(3):287–315. [aHM]
- Kuhn, D., Weinstock, M. & Flaton, R. (1994) How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science* 5(5):289–96. [aHM]
- Kuipers, B. (2000) The spatial semantic hierarchy. *Artificial Intelligence* 119 (1–2):191–233. [JW]
- Kunda, Z. (1987) Motivation and inference: Self-serving generation and evaluation of evidence. *Journal of Personality and Social Psychology* 53(4):636–47. [aHM]
- Kunda, Z. (1990) The case for motivated reasoning. *Psychological Bulletin* 108(3):480–98. [aHM]
- Lambert, A. J., Cronen, S., Chasteen, A. L. & Lickel, B. (1996) Private vs public expressions of racial prejudice. *Journal of Experimental Social Psychology* 32(5):437–59. [aHM]
- Landemore, H. (2007) *Democratic reason: Politics, collective intelligence, and the rule of the many*. Harvard University. [rHM]
- Landemore, H. (in press) Democratic reason: The mechanisms of collective intelligence in politics. In: *Collective wisdom: Principles and Mechanisms*, ed. H. Landemore & J. Elster. Cambridge University Press [aHM]
- Landemore, H. & Mercier, H. (submitted) "Talking it out": Deliberation with others versus deliberation within. [rHM]
- Langer, E. J., Blank, A. & Chanowitz, B. (1978) The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology* 36(6):635–42. [rHM]
- Lao, J. & Kuhn, D. (2002) Cognitive engagement and attitude development. *Cognitive Development* 17(2):1203–17. [aHM]
- Larson, M., Britt, M. A. & Larson, A. A. (2004) Disfluencies in comprehending argumentative texts. *Reading Psychology* 25:205–24. [rHM, CRW]
- Lassiter, G. D., Lindberg, M. J., Gonzalez-Vallejo, C., Bellezza, F. S. & Phillips, N. D. (2009) The deliberation-without-attention effect: Evidence for an artifactual interpretation. *Psychological Science* 20(6):671–75. [aHM]
- Laughlin, P. R., Bonner, B. L. & Miner, A. C. (2002) Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes* 88(2):605–20. [aHM]
- Laughlin, P. R. & Ellis, A. L. (1986) Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology* 22(3):177–89. [aHM]
- Laughlin, P. R., Hatch, E. C., Silver, J. S. & Boh, L. (2006) Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology* 90(4):644–51. [aHM]
- Laughlin, P. R., VanderStoep, S. W. & Hollingshead, A. B. (1991) Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology* 61(1):50–67. [aHM]
- Laughlin, P. R., Zander, M. L., Kniewel, E. M. & Tan, T. S. (2003) Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective reasoning. *Journal of Personality and Social Psychology* 85(4):684–94. [aHM]

- Lea, R. B. (1995) On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1469–82. [JSU]
- Lee, L., Amir, O. & Ariely, D. (2009) In search of *Homo economicus*: Preference consistency, emotions, and cognition. *Journal of Consumer Research* 36:173–87. [aHM]
- Lengwiler, M. (2008) Participatory approaches in science and technology: Historical origins and current practices in critical perspective. *Science, Technology, & Human Values* 33(2):186–200. [KC-CW]
- Lerner, J. S. & Tetlock, P. E. (1999) Accounting for the effects of accountability. *Psychological Bulletin* 125(2):255–75. [aHM]
- Leslie, A. M. (1987) Pretense and representation: The origins of a “theory of mind.” *Psychological Review* 94(4):412–26. [aHM]
- Levin, I. & Druryan, S. (1993) When sociocognitive transaction among peers fails: The case of misconceptions in science. *Child Development* 64(5):1571–91. [rHM, JEO]
- Lewontin, R. C., Rose, S. & Kamin, L. J. (1987) *Not in our genes: Biology, ideology, and human nature*. Pantheon. [DN]
- Liberman, A. & Chaiken, S. (1991) Value conflict and thought-induced attitude change. *Journal of Experimental Social Psychology* 27(3):203–16. [aHM]
- Little, B. R. (1972) Psychological man as scientist, humanist and specialist. *Journal of Experimental Research in Personality* 6:95–118. [JA]
- Littlepage, G. E. & Mueller, A. L. (1997) Recognition and utilization of expertise in problem-solving groups: Expert characteristics and behavior. *Group Dynamics* 1(4):324–28. [aHM]
- Lombardelli, C., Proudman, J. & Talbot, J. (2005) Committees versus individuals: An experimental analysis of monetary policy decision-making. *International Journal of Central Banking* 1(1):181–205. [aHM]
- Lord, C. G., Ross, L. & Lepper, M. R. (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37(11):2098–109. [aHM, JEO]
- Lucas, E. J. & Ball, L. J. (2005) Think-aloud protocols and the selection task: Evidence for relevance effects and rationalisation processes. *Thinking and Reasoning* 11(1):35–66. [aHM]
- Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., Keshavan, M. S., Genovese, C. R., Eddy, W. F. & Sweeney, J. A. (2001) Maturation of widely distributed brain function subserves cognitive development. *NeuroImage* 13(5):786–93. [DN]
- Maciejovsky, B. & Budescu, D. V. (2007) Collective induction without cooperation? Learning and knowledge transfer in cooperative groups and competitive auctions. *Journal of Personality and Social Psychology* 92(5):854–70. [aHM]
- Madsen, D. B. (1978) Issue importance and group choice shifts: A persuasive arguments approach. *Journal of Personality and Social Psychology* 36(10):1118–27. [aHM]
- Mahoney, M. J. (1977) Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1(2):161–75. [aHM]
- Mascaro, O. & Sperber, D. (2009) The moral, epistemic, and mindreading components of children’s vigilance towards deception. *Cognition* 112(3):367–80. [arHM]
- Mathews, V. P., Kronenberger, W. G., Wang, Y., Lurito, J. T., Lowe, M. J. & Dunn, D. W. (2005) Media violence exposure and frontal lobe activation measured by functional magnetic resonance imaging in aggressive and nonaggressive adolescents. *Journal of Computer Assisted Tomography* 29(3):287–92. [DN]
- Mazar, N., Amir, O. & Ariely, D. (2008) The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45(6):633–44. [aHM]
- McGuire, T. W., Kiesler, S. & Siegel, J. (1987) Group and computer-mediated discussion effects in risk decision making. *Journal of Personality and Social Psychology* 52(5):917–30. [aHM]
- McGuire, W. J. (1964) Inducing resistance to persuasion: Some contemporary approaches. In: *Advances in experimental social psychology, vol. 1*, ed. L. Berkowitz. Academic Press. [aHM]
- McKenzie, C. R. M. (2004) Framing effects in inference tasks – and why they’re normatively defensible. *Memory & Cognition* 32(6):874–85. [aHM]
- McKenzie, C. R. M. & Nelson, J. D. (2003) What a speaker’s choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review* 10(3):596–602. [aHM]
- McMackin, J. & Slovic, P. (2000) When does explicit justification impair decision making? *Journal of Applied Cognitive Psychology* 14(6):527–41. [aHM]
- Means, M. L. & Voss, J. F. (1996) Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction* 14:139–78. [CRW]
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slater, R. B. & Pennebaker, J. W. (2007) Are women really more talkative than men? *Science* 317:82. [J-LD]
- Mercier, H. (submitted a) Looking for arguments. [rHM]
- Mercier, H. (in press a) On the universality of argumentative reasoning. *Journal of Cognition and Culture*. [arHM]
- Mercier, H. (in press b) Reasoning serves argumentation in children. *Cognitive Development*. [rHM]
- Mercier, H. (in press c) What good is moral reasoning? *Mind & Society*. [rHM]
- Mercier, H. (submitted b) When experts argue: Explaining the best and the worst of reasoning. [rHM]
- Mercier, H. & Landemore, H. (in press) Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*. [arHM]
- Mercier, H. & Sperber, D. (2009) Intuitive and reflective inferences. In: *In two minds: Dual processes and beyond*, ed. J. St. B. T. Evans & K. Frankish, pp. 149–70. Oxford University Press. [arHM]
- Michaelsen, L. K., Watson, W. E. & Black, R. H. (1989) A realistic test of individual versus group consensus decision making. *Journal of Applied Psychology* 74(5):834–39. [aHM]
- Milch, K. F., Weber, E. U., Appelt, K. C., Handgraaf, M. J. J. & Krantz, D. H. (2009) From individual preference construction to group decisions: Framing effects and group processes. *Organizational Behavior and Human Decision Processes*. [aHM]
- Milford, M. (2008) *Robot navigation from nature: Simultaneous localisation, mapping, and path planning based on hippocampal models*. Springer-Verlag. [JW]
- Milford, M. & Wyeth, G. (2003) Hippocampal models for simultaneous localisation and mapping on an autonomous robot. In: *Proceedings of the Australasian Conference on Robotics and Automation, Brisbane, Australia*. Available at: <http://www.araa.asn.au/acra/acra2003/papers/35.pdf>. [JW]
- Millar, M. G. & Tesser, A. (1986) Thought-induced attitude change: The effects of schema structure and commitment. *Journal of Personality and Social Psychology* 51(2):259–69. [aHM]
- Millar, M. G. & Tesser, A. (1989) The effects of affective-cognitive consistency and thought on the attitude-behavior relation. *Journal of Experimental Social Psychology* 25(2):189–202. [aHM]
- Miller, A. G., Michoskey, J. W., Bane, C. M. & Dowd, T. G. (1993) The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology* 64(4):561–74. [aHM]
- Misak, C. (2009) Truth and democracy: Pragmatism and the deliberative virtues. In: *Does truth matter? Democracy and public space*, ed. R. Geenens & R. Tinnevelt, pp. 29–39. Springer. [KC-CW]
- Molden, D. C. & Higgins, E. T. (2005) Motivated thinking. In: *The Cambridge handbook of thinking and reasoning*, ed. K. Holyoak & R. Morrison. Cambridge University Press. [aHM]
- Moore, A. B., Clark, B. A. & Kane, M. J. (2008) Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science* 19(6):549–57. [aHM]
- Moorhead, G., Ferench, R. & Neck, C. P. (1991) Group decision fiascoes continue: Space shuttle Challenger and a revised groupthink framework. *Human Relations* 44(6):539–50. [aHM]
- Morgan-Olsen, B. (2010) Conceptual exclusion and public reason. *Philosophy of the Social Sciences* 40(2):213–43. [KC-CW]
- Morris, B. & Hasson, U. (2010) Multiple sources of competence underlying the comprehension of inconsistencies: A developmental investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36:277–87. [JEO]
- Morsanyi, K. & Handley, S. J. (2008) How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology* 99(1):18–36. [aHM]
- Moscovici, S. & Zavalloni, M. (1969) The group as a polarizer of attitudes. *Journal of Personality and Social Psychology* 12:125–35. [RJS]
- Moser, E. I., Kropff, E. & Moser, M. B. (2008) Place cells, grid cells, and the brain’s spatial representation system. *Annual Review of Neuroscience* 31:69–89. [JW]
- Moshman, D. & Geil, M. (1998) Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning* 4(3):231–48. [aHM]
- Narvaez, D. (2008) Triune ethics: The neurobiological roots of our multiple moralities. *New Ideas in Psychology* 26:95–119. [DN]
- Narvaez, D. (2010) Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science* 5(2):163–81. [DN]
- Navarro, A. D. & Fantino, E. (2005) The sunk cost effect in pigeons and humans. *Journal of the Experimental Analysis of Behavior* 83(1):1–13. [aHM]
- Neuman, Y. (2003) Go ahead, prove that God does not exist! On high school students’ ability to deal with fallacious arguments. *Learning and Instruction* 13(4):367–80. [aHM]
- Neuman, Y., Weinstock, M. P. & Glasner, A. (2006) The effect of contextual factors on the judgment of informal reasoning fallacies. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 59:411–25. [MH, arHM]

- Newell, A. & Simon, H. A. (1956) The logic theory machine: A complex information processing system. *IRE Transactions on Information Theory* IT-2(3):61–79. [JW]
- Newell, B. R., Wong, K. Y., Cheung, J. C. H. & Rakow, T. (2009) Think, blink or sleep on it? The impact of modes of thought on complex decision making. *Quarterly Journal of Experimental Psychology* 62(4):707–32. [aHM]
- Newstead, S. E., Handley, S. J. & Buck, E. (1999) Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory & Cognition* 27(2):344–54. [aHM]
- Nickerson, R. S. (1998) Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology* 2(2):175–220. [aHM, E]NS]
- Niv, Y. & Schoenbaum, G. (2008) Dialogues on prediction errors. *Trends in Cognitive Sciences* 12(7):265–72. [aHM]
- Norenzayan, A., Smith, E. E., Kim, B. J. & Nisbett, R. E. (2002) Cultural preferences for formal versus intuitive reasoning. *Cognitive Science* 26(5):653–84. [rHM]
- Novaes, C. D. (2005) Medieval *obligationes* as logical games of consistency maintenance. *Synthese* 145(3):371–95. [aHM]
- Nussbaum, E. M. (2008) Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology* 33(3):345–59. [aHM]
- Nussbaum, E. M. & Sinatra, G. M. (2003) Argument and conceptual engagement. *Contemporary Educational Psychology* 28(3):384–95. [aHM]
- O’Keefe, J. & Dostrovsky, J. (1971) The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* 34(1):171–75. [JW]
- O’Keefe, J. & Nadel, L. (1978) *The hippocampus as a cognitive map*. Oxford University Press. [JW]
- Oaksford, M. & Chater, N. (2007) *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. [aHM, MO]
- Oaksford, M. & Chater, N. (2009) The uncertain reasoner: Bayes, logic, and rationality. *Behavioral and Brain Sciences* 32:105–20. [JF]
- Oaksford, M., Chater, N. & Grainger, R. (1999) Probabilistic effects in data selection. *Thinking & Reasoning* 5(3):193–243. [aHM]
- Oaksford, M., Chater, N. & Larkin, J. (2000) Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26:883–99. [MO]
- Oaksford, M. & Hahn, U. (2004) A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology* 58(2):75–85. [aHM]
- Oaksford, M. & Hahn, U. (2007) Induction, deduction and argument strength in human reasoning and argumentation. In *Inductive reasoning*, ed. A. Feeney & E. Heit, pp. 269–301. Cambridge University Press. [MO]
- Okada, E. M. (2005) Justification effects on consumer choice of hedonic and utilitarian goods. *Journal of Marketing Research* 42(1):43–53. [aHM]
- Okada, T. & Simon, H. A. (1997) Collaboration discovery in a scientific domain. *Cognitive Science* 21(2):109–46. [aHM]
- OpenClinical. (2001–2006) *PROforma: Formal knowledge representation method for the development and execution of clinical guidelines*. Available at: [http://www.openclinical.org/gmm\\_proforma.html](http://www.openclinical.org/gmm_proforma.html). [JF]
- Ormerod, P. (2005) *Why most things fail: Evolution, extinction and economics*. Faber & Faber. [aHM]
- Paese, P. W., Bieser, M. & Tubbs, M. E. (1993) Framing effects and choice shifts in group decision making. *Organizational Behavior and Human Decision Processes* 56(1):149–56. [aHM]
- Peirce, C. S. (1931–35) *Collected papers of Charles Sanders Peirce*. Harvard University Press. [JW]
- Pennington, N. & Hastie, R. (1993) Reasoning in explanation-based decision-making. *Cognition* 49(1–2):123–63. [aHM]
- Perelman, C. & Olbrechts-Tyteca, L. (1969) *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press. [aHM, MO]
- Perkins, D. N. (1985) Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology* 77(5):562–71. [aHM]
- Perkins, D. N., Allen, R. & Hafner, J. (1983) Difficulties in everyday reasoning. In: *Thinking: The expanding frontier*, ed. W. Maxwell, pp. 177–89. Franklin Institute Press. [CRW]
- Perkins, D. N., Farady, M. & Bushey, B. (1991) Everyday reasoning and the roots of intelligence. In *Informal reasoning*, ed. J. Voss, D. N. Perkins & J. Segal, pp. 83–105. Erlbaum. [CRW]
- Petersen, M. B., Sell, A., Tooby, J. & Cosmides, L. (2010) Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. In: *Human Morality and Sociality*, ed. H. Høgh-Olesen, pp. 72–131. Palgrave Macmillan. [DP]
- Petty, R. E. & Cacioppo, J. T. (1979) Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology* 37(10):1915–26. [arHM, JSU]
- Petty, R. E. & Cacioppo, J. T. (1996) *Attitudes and persuasion: Classic and contemporary approaches*. Westview Press. [JEO]
- Petty, R., Cacioppo, J. & Goldman, R. (1981) Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology* 41(5):847–55. [rHM]
- Petty, R. E. & Wegener, D. T. (1998) Attitude change: Multiple roles for persuasion variables. In: *The handbook of social psychology*, vol. 1, ed. D. Gilbert, S. Fiske & G. Lindzey, pp. 323–90. McGraw-Hill. [aHM]
- Pietraszewski, D., Curry, O., Petersen, M. B. & Tooby, J. (in preparation) Politics erases race but not sex: Evidence that signals of political party support engage coalitional psychology. [DP]
- Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13(4):707–84. [rHM]
- Poletiek, F. H. (1996) Paradoxes of falsification. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 49(2):447–62. [aHM, FHP]
- Poletiek, F. H. (2001) *Hypothesis-testing behavior*. Essays in Cognitive Psychology Series. Psychology Press. [FHP]
- Poletiek, F. H. & Bermsden, M. (2000) Hypothesis testing as risk behavior with regard to beliefs. *Journal of Behavioral Decision Making* 13:107–23. [FHP]
- Pomerantz, E. M., Chaiken, S. & Tordesillas, R. S. (1995) Attitude strength and resistance processes. *Journal of Personality and Social Psychology* 69(3):408–19. [aHM]
- Powell, C. (2003) The Delphi technique: Myths and realities. *Journal of Advanced Nursing* 41(4):376–82. [aHM]
- Prasad, M., Perrin, A. J., Bezila, K., Hoffman, S. G., Kindleberger, K., Manturuk, K. & Powers, A. (2009) “There must be a reason”: Osama, Saddam, and inferred justification. *Sociological Inquiry* 79(2):142–62. [aHM]
- Premack, D. & Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1(4):515–26. [aHM]
- Pritchard, D. (2005) *Epistemic luck*. Clarendon Press. [aHM]
- Pyszczynski, T. & Greenberg, J. (1987) Toward and integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In: *Advances in experimental social psychology*, vol. 20, ed. L. Berkowitz, pp. 297–340. Academic Press. [aHM]
- Quayle, J. D. & Ball, L. J. (2000) Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 53:1202–223. [E]NS]
- Rader, A. W. & Sloutsky, V. M. (2002) Processing of logically valid and logically invalid conditional inferences in discourse comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:59–68. [JSU]
- Ratneshwar, S., Shocker, A. D. & Stewart, D. W. (1987) Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research* 13(4):520–33. [aHM]
- Reb, J. (2005) The role of regret aversion in decision making. Unpublished doctoral dissertation, University of Arizona, Tucson. [TC]
- Reb, J. (2008) Regret aversion and decision process quality: Effects of regret salience on decision process carefulness. *Organizational Behavior and Human Decision Processes* 105:169–82. [TC]
- Reb, J. & Connolly, T. (2010) The effects of action, normality, and decision carefulness on anticipated regret: Evidence for a broad meditating role of decision justifiability. *Cognition & Emotion* 24:1405–20. [TC]
- Reb, J. & Connolly, T. (2009) Myopic regret avoidance in repeated decision making. *Organizational Behavior and Human Decision Processes* 109:182–89. [TC]
- Recanati, F. (2000) *Oratio obliqua, oratio recta*. MIT Press. [aHM]
- Redlawsk, D. P. (2002) Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *Journal of Politics* 64(4):1021–44. [aHM]
- Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H. & Holowchak, M. (1993) Reasoning in conversation. *Cognition and Instruction* 11(3–4):347–64. [aHM]
- Ricco, R. B. (2003) The macrostructure of informal arguments: A proposed model and analysis. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 56(6):1021–51. [aHM, MH]
- Rim, S., Min, K. E., Uleman, J. S. & Chartrand, T. L. (2010) A functional analysis of stages of spontaneous impression formation, serving nonconscious affiliation goals. Unpublished manuscript. [JSU]
- Rim, S., Uleman, J. S. & Trope, Y. (2009) Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology* 45:1088–97. [JSU]
- Rips, L. J. (1994) *The psychology of proof: Deductive reasoning in human thinking*. MIT Press. [arHM]
- Rips, L. J. (1998) Reasoning and conversation. *Psychological Review* 105(3):411–41. [aHM]
- Rips, L. J. (2002) Circular reasoning. *Cognitive Science* 26(6):767–95. [MH, arHM]
- Ritchart, R. & Perkins, D. N. (2005) Learning to think: The challenges of teaching thinking. In: *The Cambridge handbook of thinking and reasoning*, ed. K. Holyoak & R. Morrison. Cambridge University Press. [aHM]
- Roberts, M. J. & Newton, E. J. (2001) Inspection times, the change task, and the rapid response selection task. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 54(4):1031–48. [aHM]

- Ross, L., Lepper, M. R. & Hubbard, M. (1975) Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology* 32(5):880–92. [aHM]
- Ross, M., McFarland, C. & Fletcher, G. J. (1981) The effect of attitude on the recall of personal histories. *Journal of Personality and Social Psychology* 40(4):627–34. [aHM]
- Rowe, G. & Wright, G. (1999) The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting* 15(4):353–75. [aHM]
- Rozin, P., Millman, L. & Nemeroff, C. (1986) Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology* 50(4):703–12. [aHM]
- Russo, J. E., Carlson, K. A. & Meloy, M. G. (2006) Choosing an inferior alternative. *Psychological Science* 17(10):899–904. [aHM]
- Ryan, W. (1971) *Blaming the victim*. Pantheon. [aHM]
- Ryfe, D. M. (2005) Does deliberative democracy work? *Annual Review of Political Science* 8:49–71. [KC-CW]
- Sá, W. C., Kelley, C. N., Ho, C. & Stanovich, K. E. (2005) Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences* 38(5):1149–61. [MH, arHM]
- Sacco, K. & Bucciarelli, M. (2008) The role of cognitive and socio-cognitive conflict in learning to reason. *Mind & Society* 7(1):1–19. [aHM]
- Sadler, O. & Tesser, A. (1973) Some effects of salience and time upon interpersonal hostility and attraction during social isolation. *Sociometry* 36(1):99–112. [aHM]
- Sahlins, M. (2008) *The Western illusion of human nature*. Prickly Paradigm Press. [DN]
- Sanitioso, R., Kunda, Z. & Fong, G. T. (1990) Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology* 59(2):229–41. [aHM]
- Savage, L. J. (1954) *The foundations of statistics*. Wiley. [aHM]
- Schaie, K. W. & Willis, S. L. (2010) *Handbook of the psychology of aging*, 7th edition. Academic Press. [DN]
- Scheibehenne, B., Greifeneder, R. & Todd, P. M. (2009) What moderates the too-much-choice effect? *Psychology & Marketing* 26(3):229–53. [aHM]
- Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R. & Frey, D. (2006) Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology* 91(6):1080–93. [aHM]
- Schweitzer, M. E. & Hsee, C. K. (2002) Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty* 25(2):185–201. [aHM]
- Sela, A., Berger, J. & Liu, W. (2009) Variety, vice, and virtue: How assortment size influences option choice. *Journal of Consumer Research* 35(6): 941–51. [aHM]
- Sell, A. (2006) Regulating welfare tradeoff ratios: Three tests of an evolutionary-computational model of human anger. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 66(8-B):4516. [rHM, DP]
- Sengupta, J. & Fitzsimons, G. J. (2000) The effects of analyzing reasons for brand preferences: Disruption or reinforcement? *Journal of Marketing Research* 37(3):318–30. [aHM]
- Sengupta, J. & Fitzsimons, G. J. (2004) The effect of analyzing reasons on the stability of brand attitudes: A reconciliation of opposing predictions. *Journal of Consumer Research* 31(3):705–11. [aHM]
- Shafir, E. & Tversky, A. (1992) Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology* 24(4):449–74. [TC, aHM]
- Shafir, E., Simonson, I. & Tversky, A. (1993) Reason-based choice. *Cognition* 49(1–2):11–36. [aHM]
- Shaw, V. F. (1996) The cognitive processes in informal reasoning. *Thinking & Reasoning* 2:51–80. [MH, aHM]
- Simon, H. A. (1955) A behavioral model of rational choice. *Quarterly Journal of Economics* 69(1):99–118. [aHM]
- Simonson, I. (1989) Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research* 16(2):158–74. [TC, aHM]
- Simonson, I. (1990) The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research* 27(2):150–62. [aHM]
- Simonson, I., Carmon, Z. & O'Curry, S. (1994) Experimental evidence on the negative effect of product features and sales promotions on brand choice. *Marketing Science* 13(1):23–40. [aHM]
- Simonson, I. & Nowlis, S. M. (2000) The role of explanations and need for uniqueness in consumer decision making: Unconventional choices based on reasons. *Journal of Consumer Research* 27(1):49–68. [aHM]
- Simonson, I., Nowlis, S. M. & Simonson, Y. (1993) The effect of irrelevant preference arguments on consumer choice. *Journal of Consumer Psychology* 2(3):287–306. [aHM]
- Simonson, I. & Nye, P. (1992) The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes* 51:416–46. [TC, aHM]
- Skowronski, J. J., Carlston, D. E., Mae, L. & Crawford, M. T. (1998) Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology* 74:837–48. [JSU]
- Slaughter, J. E., Bagger, J. & Li, A. (2006) Context effects on group-based employee selection decisions. *Organizational Behavior and Human Decision Processes* 100: 47–59. [TC]
- Slavin, R. E. (1995) *Cooperative learning: Theory, research and practice*, 2nd ed. Allyn & Bacon. [aHM]
- Slovan, S. A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119(1):3–22. [aHM]
- Slovic, P. (1975) Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception and Performance* 1:280–87. [TC, aHM]
- Smith, J. A., Weber, E. U., Appelt, K. C. & Milch, K. F. (2009) Querying the group mind: Applying query theory to group discussions. Poster presented at the Annual Meeting of the Society for Judgment and Decision Making, Boston. [EUW]
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N. & Su, T. T. (2009) Why peer discussion improves student performance on in-class concept questions. *Science* 323(5910):122–24. [aHM]
- 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):464–93. [aHM]
- Snizek, J. A. & Henry, R. A. (1989) Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes* 43(1):1–28. [aHM]
- Snyder, M., Kleck, R. E., Strenta, A. & Mentzer, S. J. (1979) Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology* 37(12):2297–306. [aHM]
- Soman, D. & Cheema, A. (2001) The effect of windfall gains on the sunk-cost effect. *Marketing Letters* 12(1):51–62. [aHM]
- Spelke, E. S. & Kinzler, K. D. (2007) Core knowledge. *Developmental Science* 10(1):89–96. [aHM]
- Sperber, D. (1997) Intuitive and reflective beliefs. *Mind and Language* 12(1):67–83. [aHM]
- Sperber, D. (2000a) Metarepresentations in an evolutionary perspective. In: *Metarepresentations: A multidisciplinary perspective*, ed. D. Sperber, pp. 117–37. Oxford University Press. [aHM]
- Sperber, D., ed. (2000b) *Metarepresentations: A multidisciplinary perspective*. Oxford University Press. [aHM]
- Sperber, D. (2001) An evolutionary perspective on testimony and argumentation. *Philosophical Topics* 29:401–13. [aHM]
- Sperber, D. (2009) L'effet gourou. *L'autre côté* 1:17–23. [rHM]
- Sperber, D., Cara, F. & Girotto, V. (1995) Relevance theory explains the selection task. *Cognition* 57(1):31–95. [aHM]
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G. & Wilson, D. (2010) Epistemic vigilance. *Mind & Language* 25(4):359–93. [arHM]
- Sperber, D. & Wilson, D. (1995) *Relevance: Communication and cognition*, 2nd ed. Blackwell. [rHM]
- Sperber, D. & Wilson, D. (2002) Pragmatics, modularity and mind-reading. *Mind and Language* 17(1–2):3–23. [aHM]
- Stanovich, K. E. (1993) Dysrationalia: A new specific learning disability. *Journal of Learning Disabilities* 26(8):501–15. [RJS]
- Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning*. Erlbaum. [JStBTE, KF]
- Stanovich, K. E. (2004) *The robot's rebellion: Finding meaning the age of Darwin*. Chicago University Press. [JStBTE, aHM]
- Stanovich, K. E. (2009) *What intelligence tests miss: The psychology of rational thought*. Yale University Press. [JStBTE, RJS]
- Stanovich, K. E. (2010) *Rationality and the reflective mind*. Oxford University Press. [JStBTE]
- Stanovich, K. E. & West, R. F. (1998) Individual differences in rational thought. *Journal of Experimental Psychology: General* 127(2):161–88. [aHM]
- Stanovich, K. E. & West, R. F. (2000) Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23:645–65. [EJNS]
- Stanovich, K. E. & West, R. F. (2003) Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In: *Evolution and the psychology of thinking*, ed. D. E. Over, pp. 171–230. Psychology Press. [JStBTE]
- Stanovich, K. E. & West, R. F. (2007) Natural myside bias is independent of cognitive ability. *Thinking & Reasoning* 13(3):225–47. [aHM]
- Stanovich, K. E. & West, R. F. (2008a) On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning* 14(2):129–67. [aHM]
- Stanovich, K. E. & West, R. F. (2008b) On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology* 94(4):672–95. [aHM]

- Stasson, M. F., Kameda, T., Parks, C. D., Zimmerman, S. K. & Davis, J. H. (1991) Effects of assigned group consensus requirement on group problem solving and group members' learning. *Social Psychology Quarterly* 54(1):25–35. [aHM]
- Staw, B. M. (1981) The escalation of commitment to a course of action. *Academy of Management Review* 6(4):577–87. [aHM]
- Stein, N. L., Bernas, R. S. & Calicchia, D. J. (1997) Conflict talk: Understanding and resolving arguments. In: *Conversation: Cognitive, communicative, and social perspectives*, ed. T. Givón, pp. 233–68. John Benjamins. [aHM]
- Stein, N. L., Bernas, R. S., Calicchia, D. J. & Wright, A. (1996) Understanding and resolving arguments: The dynamics of negotiation. In: *Models of understanding text*, ed. B. Britton & A. G. Graesser, pp. 257–88. Erlbaum. [aHM]
- Steiner, I. D. (1972) *Group processes and productivity*. Academic Press. [aHM]
- Sterelny, K. (in press) *The evolved apprentice*. MIT Press. [aHM]
- Sternberg, R. J., ed. (2002) *Why smart people can be so stupid*. Yale University Press. [RJS]
- Sternberg, R. J., Reznitskaya, A. & Jarvin, L. (2007) Teaching for wisdom: What matters is not just what students know, but how they use it. *London Review of Education* 5(2):143–58. [RJS]
- Stupple, E. J. N. & Ball, L. J. (2008) Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning* 14:168–81. [EJNS]
- Sunstein, C. R. (2002) The law of group polarization. *Journal of Political Philosophy* 10(2):175–95. [aHM]
- Swinney, D. (1979) Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18:645–60. [JSU]
- Taber, C. S. & Lodge, M. (2006) Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50(3):755–69. [aHM]
- Taleb, N. N. (2007) *The black swan: The impact of the highly improbable*. Random House. [aHM]
- Talisie, R. B. (2009) Folk epistemology and the justification of democracy. In: *Does truth matter? Democracy and public space*, ed. R. Geenens & R. Tinnevelt, pp. 41–54. Springer. [KC-CW]
- Taube, J. S., Muller, R. U. & Ranck, J. B., Jr. (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience* 10(2):420–35. [JW]
- Tesser, A. (1976) Attitude polarization as a function of thought and reality constraints. *Journal of Research in Personality* 10(2):183–94. [aHM]
- Tesser, A. & Conlee, M. C. (1975) Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology* 31(2):262–70. [aHM]
- Tesser, A. & Leone, C. (1977) Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology* 13(4):340–56. [aHM]
- Tetlock, P. E. (1998) Close-call counterfactuals and belief-system defenses: I was not almost wrong but I was almost right. *Journal of Personality and Social Psychology* 75(3):639–52. [aHM]
- Tetlock, P. E. & Boettger, R. (1989) Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology* 57(3):388–98. [aHM]
- Tetlock, P. E., Lerner, J. S. & Boettger, R. (1996) The dilution effect: Judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology* 26(6):915–34. [aHM]
- Tetlock, P. E., Skitka, L. & Boettger, R. (1989) Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology* 57(4):632–40. [aHM]
- Thoma, S. J. & Bebeau, M. (2008) *Moral Judgment competency is declining over time: Evidence from 20 years of defining issues test data*. Paper presented to the American Educational Research Association, New York. [DN]
- Thompson, D. F. (2008) Deliberative democratic theory and empirical political science. *Annual Review of Political Science* 11:497–520. [KC-CW]
- Thompson, D. V., Hamilton, R. W. & Rust, R. T. (2005a) Feature fatigue: When product capabilities become too much of a good thing. *Journal of Marketing Research* 42(4):431–42. [aHM]
- Thompson, D. V. & Norton, M. I. (2008) The social utility of feature creep. In: *Advances in consumer research*, vol. 35, ed. A. Lee & D. Soman, pp. 181–84. Association for Consumer Research. [aHM]
- Thompson, V. A., Evans, J. St. B. T. & Handley, S. J. (2005b) Persuading and dissuading by conditional argument. *Journal of Memory and Language* 53(2):238–57. [MH, arHM]
- Thompson, V. A., Newstead, S. E. & Morley, N. J. (2010) Methodological and theoretical issues in belief-bias: Implications for dual process theories. In: *The science of reason: A festschrift for Jonathan St. B. T. Evans*, ed. K. I. Manktelow, D. E. Over & S. Elqayam, pp. 309–37. Psychology Press. [EJNS]
- Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W. & Campbell, J. I. D. (2003) Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review* 10(1):184–89. [aHM, EJNS]
- Thorsteinson, T. J. & Withrow, S. (2009) Does unconscious thought outperform conscious thought on complex decisions? A further examination. *Judgment and Decision Making* 4(3):235–47. [aHM]
- Thrun, S. (2003) Robotic mapping: A survey. In: *Exploring artificial intelligence in the new millennium*, ed. G. Lakemeyer & B. Nebel, pp. 1–36. Morgan Kaufmann. [JW]
- Tichy, G. (2004) The over-optimism among experts in assessment and foresight. *Technological Forecasting & Social Change* 71(4):341–63. [aHM]
- Tindale, R. S. & Sheffield, S. (2002) Shared information, cognitive load, and group memory. *Group Processes & Intergroup Relations* 5(1):5–18. [aHM]
- Todorov, A. & Uleman, J. S. (2002) Spontaneous trait inferences are bound to actors: Evidence from false recognition. *Journal of Personality and Social Psychology* 83:1051–65. [JSU]
- Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review* 55(4):189–208. [JW]
- Tolmie, A., Howe, C., Mackenzie, M. & Greer, K. (1993) Task design as an influence on dialogue and learning: Primary school group work with object flotation. *Social Development* 2(3):183–201. [aHM]
- Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28(5):675–91. [aHM]
- Tomkins, S. S. (1965) Affect and the psychology of knowledge. In: *Affect, cognition, and personality: Empirical studies*, ed. S. S. Tomkins & C. E. Izard, pp. 72–97. Springer. [JA]
- Tooby, J., Cosmides, L., Sell, A., Lieberman, D. & Szycer, D. (2008) Internal regulatory variables and the design of human motivation: A computational and evolutionary approach. In: *Handbook of approach and avoidance motivation*, ed. A. Elliot, pp. 251–71. Psychology Press. [DP]
- Toplak, M. E. & Stanovich, K. E. (2003) Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology* 17:851–60. [CRW]
- Trogon, A. (1993) How does the process of interaction work when two interlocutors try to resolve a logical problem? *Cognition and Instruction* 11(3–4):325–45. [aHM]
- Tversky, A. & Kahneman, D. (1981) The framing of decisions and the psychology of choice. *Science* 211(4481):453–58. [aHM]
- Tversky, A. & Kahneman, D. (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90(4):293–315. [aHM]
- Tversky, A. & Shafir, E. (1992) The disjunction effect in choice under uncertainty. *Psychological Science* 3(5):305–309. [aHM]
- Tversky, A., Sattath, S. & Slovic, P. (1988) Contingent weighting in judgment and choice. *Psychological Review* 95(3):371–84. [aHM]
- Twardy, C. R. (2004) Argument maps improve critical thinking. *Teaching Philosophy* 27:95–116. [MH]
- Twene, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A. & Arkkelin, D. L. (1980) Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology* 32(1):109–23. [aHM]
- Twenge, J. & Campbell, R. (2009) *The narcissism epidemic: Living in the age of entitlement*. Free Press. [DN]
- Uleman, J. S. (2005) On the inherent ambiguity of traits and other mental concepts. In: *Other minds: How humans bridge the divide between self and others*, ed. B. F. Malle & S. D. Hodges, pp. 253–67. Guilford. [JSU]
- Uleman, J. S., Newman, L. S. & Moskowitz, G. B. (1996) People as flexible interpreters: evidence and issues from spontaneous trait inference. In: *Advances in experimental social psychology*, vol. 28, ed. M. P. Zanna, pp. 211–79. Academic Press. [JSU]
- Uleman, J. S., Saribay, S. A. & Gonzalez, C. (2008) Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology* 59:329–60. [JSU]
- Valdesolo, P. & DeSteno, D. (2008) The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology* 44(5):1334–38. [aHM]
- van Boxtel, C., van der Linden, J. & Kanselaar, G. (2000) Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction* 10(4):311–30. [aHM]
- Van Gelder, T. (2005) Teaching critical thinking: Some lessons from cognitive science. *College Teaching* 53:41–46. [MH]
- Van Gelder, T., Bissett, M. & Cumming, G. (2004) Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology* 58:142–52. [MH]
- Vinokur, A. (1971) Review and theoretical analysis of the effects of group processes upon individual and group decisions involving risk. *Psychological Bulletin* 76(4):231–50. [aHM]
- Vinokur, A. & Burnstein, E. (1978) Depolarization of attitudes in groups. *Journal of Personality and Social Psychology* 36(8):872–85. [aHM]
- Voss, J. F. & Van Dyke, J. A. (2001) Argumentation in psychology: Background comments. *Discourse Processes* 32:89–111. [CRW]
- Vygotsky, L. (1978) *Mind in society: The development of higher psychological processes*. Harvard University Press. [DK]

- Vygotsky, L. (1986) *Thought and language*. Trans. A. Kozulin. MIT Press. (Original work published 1934.) [PG-S]
- Wason, P. C. (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 12(3):129–37. [aHM, FHP]
- Wason, P. C. (1966) Reasoning. In: *New horizons in psychology: I*, ed. B. M. Foss, pp. 106–37. Penguin. [aHM]
- Wason, P. C. & Evans, J. St. B. T. (1975) Dual processes in reasoning? *Cognition* 3:141–54. [JStBTE, aHM]
- Webb, N. M. & Palinscar, A. S. (1996) Group processes in the classroom. In: *Handbook of educational psychology*, ed. D. C. Berliner & R. C. Calfee, pp. 841–73. Prentice-Hall. [aHM]
- Weber, E. U. & Johnson, E. J. (2006) Constructing preferences from memory. In: *The construction of preference*, ed. S. Lichtenstein & P. Slovic, pp. 397–410. Cambridge University Press. [EUW]
- Weber, E. U. & Johnson, E. J. (2009) Mindful judgment and decision making. *Annual Review of Psychology* 60:53–86. [EUW]
- Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. & Goldstein, D. G. (2007) Asymmetric discounting in intertemporal choice: A query theory account. *Psychological Science* 18(6):516–23. [aHM, EUW]
- Weinstock, M., Neuman, Y. & Tabak, I. (2004) Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology* 29(1):77–94. [MH, aHM]
- Wetherick, N. E. (1962) Eliminative and enumerative behavior in a conceptual task. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 14:246–49. [FHP]
- Whiten, A. & Byrne, R. W., eds. (1997) *Machiavellian intelligence II: Extensions and evaluations*. Cambridge University Press. [aHM]
- Willingham, D. T. (2008) Critical thinking: Why is it so hard to teach? *Arts Education Policy Review* 109(4):21–32. [aHM]
- Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B. & Rotondo, J. A. (1984) Effects of analyzing reasons on attitude-behavior consistency. *Journal of Personality and Social Psychology* 47(1):5–16. [aHM]
- Wilson, T. D., Dunn, D. S., Kraft, D. & Lisle, D. J. (1989a) Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In: *Advances in experimental social psychology, vol. 19*, ed. L. Berkowitz, pp. 123–205. Academic Press. [aHM]
- Wilson, T. D., Kraft, D. & Dunn, D. S. (1989b) The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology* 25(5):379–400. [aHM]
- Wilson, T. D. & LaFleur, S. J. (1995) Knowing what you'll do: Effects of analyzing reasons on self-prediction. *Journal of Personality and Social Psychology* 68(1):21–35. [aHM]
- Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J. & LaFleur, S. J. (1993) Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin* 19(3):331–39. [aHM]
- Wilson, T. D. & Schooler, J. W. (1991) Thinking too much: Introspection can reduce the quality of preferences and decisions. *Thinking* 60(2):181–92. [aHM]
- Wolfe, C. R., Albrecht, M. J. & Britt, M. A. (2007) Any reason is better than none: Implausible reasons in argumentation. Paper presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA. [CRW]
- Wolfe, C. R. & Boone, W. J. (under review) Individual differences in the “MySide bias” in reasoning and argumentation. [CRW]
- Wolfe, C. R. & Britt, M. A. (2005) The use of other side information: Explaining the myside bias in argumentation. Paper presented at the 46th Annual Meeting of the Psychonomic Society, Toronto, Canada. [CRW]
- Wolfe, C. R. & Britt, M. A. (2008) Locus of the my-side bias in written argumentation. *Thinking & Reasoning* 14(1):1–27. [rHM, CRW]
- Wolfe, C. R. & Britt, M. A. (2009) Individual differences in the “myside bias” in reasoning and argumentation. Paper presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA. [CRW]
- Wolfe, C. R., Britt, M. A. & Butler, J. A. (2009a) Argumentation schema and the myside bias in written argumentation. *Written Communication* 26:183–209. [CRW]
- Wolfe, C. R., Britt, M. A., Petrovic, M., Albrecht, M. & Kopp, K. (2009b) The efficacy of a Web-based counterargument tutor. *Behavior Research Methods* 41:691–98. [CRW]
- Wolpert, D. M. & Kawato, M. (1998) Multiple paired forward and inverse models for motor control. *Neural Networks* 11(7–8):1317–29. [aHM]
- Wu, K. C.-C. (2008) Expanding the vision of visual bioethics. *American Journal of Bioethics* 8(12):63–64. [KC-CW]
- Wynne, B. (1996) May the sheep safely graze? A reflexive view of the expert–lay knowledge divide. In: *Risk, environment & modernity: Towards a new ecology*, ed. S. Lash, B. Szerszynski & B. Wynne, pp. 44–83. Sage. [KC-CW]
- Xu, J. & Schwarz, N. (2009) Do we really need a reason to indulge? *Journal of Marketing Research* 46(1):25–36. [aHM]
- Yates, J. F., Lee, J.-W. & Shinotsuka, H. (1992) Cross-national variation in probability judgment. Paper presented at the Annual Meeting of the Psychonomic Society, St. Louis. [aHM]
- Yegnashankaran, K. (2010) Reasoning as action. Unpublished doctoral dissertation, Harvard University. [PG-S]
- Zahavi, A. & Zahavi, A. (1997) *The handicap principle: A missing piece of Darwin's puzzle*. Oxford University Press. [aHM]