SUBMITTED VERSION – NOT TO BE QUOTED

Why do humans reason?

Arguments for an argumentative theory

Hugo Mercier

Philosophy, Politics and Economics Program

University of Pennsylvania

313 Cohen Hall

249 South 36th Street

Philadelphia, PA 19104

hmercier@sas.upenn.edu

http://hugo.mercier.googlepages.com/

Dan Sperber

Institut Jean Nicod (EHESS-ENS-CNRS)

29, rue d'Ulm

75005, Paris

dan@sperber.com

http://www.dan.sperber.com

Keywords.
Argumentation, Confirmation bias, Decision making, Dual process theory, Evolutionary psychology, Motivated reasoning, Reason-based choice, Reasoning.

Short abstract (95 words).
Reasoning is generally seen as a mean to improve knowledge and make better decisions. Much evidence, however, shows that reasoning often leads to epistemic distortions and poor decisions. This suggests rethinking the function of reasoning. 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 human exceptional dependence on communication and vulnerability to misinformation. A wide range of evidence in the psychology or reasoning and decision making can be reinterpreted and better explained in the light of this hypothesis.

Long abstract (245 words).
Reasoning is generally seen as a mean to improve knowledge and make better decisions. Much evidence, however, shows that reasoning often leads to epistemic distortions and poor decisions. This suggests rethinking the function of reasoning. 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 human exceptional dependence on communication and vulnerability to misinformation. A wide range of evidence in the psychology or 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 with the perspective of having to defend their opinions. Reasoning so motivated can distort evaluations and attitudes and allow the persistence of erroneous beliefs. Proactively used reasoning also favors decisions that are easy to justify but not necessarily better. In all of 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 favor conclusions in support of which arguments can be found.

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 expectation 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, Mamassian, & Yuille, 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 a new mental representation (or 'conclusion') is consciously produced, but also the previously held representations that warrant it (or 'premises') are consciously entertained. The premises are seen as providing reasons to accept the conclusion. Most work in the psychology of reasoning is about such conscious mental processes. Conscious reasoning is typically human. There is no evidence that it occurs in non-human animals or in pre-verbal children.[i]

How do humans reason? Why do they reason? These two questions are mutually relevant since the mechanisms of reasoning are likely to be adjusted to its function. While the how-question has been systematically investigated (e.g., Evans, Newstead, & Byrne, 1993; Johnson-Laird, 2006; 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 attention. According to a long philosophical tradition, reasoning allows the human mind to go beyond mere perception, habit, and instinct; human rationality is seen as grounded in the capacity to reason and as what separates humans from beasts. Here, after a brief discussion of the mechanisms of inference in general and of reasoning in particular, we focus on the why-question and propose that the primary function of reasoning is to permit the production and evaluation of arguments in communication. This understanding of the function of reasoning helps, we argue, explain major features of the reasoning process.

1.1 Intuitive and reflective inference

Since the 1960s, much work in the psychology or 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, Slovic, & Tversky, 1982). This work has led to a rethinking of the mechanisms of reasoning, but not—or at least not to the same degree—of its assumed function of enhancing human cognition and decision making.

What has been questioned is the view that reasoning is the outcome of a single, unitary reasoning faculty. The most radical challenge has come from work by Leda Cosmides, John Tooby and their collaborators on evolved psychological adaptations (Cosmides, 1989; Fiddick, Cosmides, & Tooby, 2000) and related work by Gerd Gigerenzer and his collaborators on bounded rationality (Gigerenzer, Todd, & ABC Research Group, 1999). These authors argue that humans are not endowed with a single domain-general reasoning ability but with a variety of domain-specific heuristic mechanisms that are tailored to solve specific problems and that take advantage of ecological regularities in the particular domain they reason about. From such a viewpoint, the function of these mechanisms--which are 'reasoning' mechanisms only in a wide sense of the term--is not to enhance cognition in general but to boost it in a variety of specific domains corresponding in particular to problems met in the ancestral environment of the species.

A less radical and more widespread challenge to a unitary view of reasoning mechanisms favours a dual-process or dual-system approach. This approach had been developed in other areas of psychology (Evans & Frankish, 2009) and in particular in social psychology (Chaiken & Trope, 1999; Wilson, Lindsey, & Schooler, 2000), where it had been first applied to persuasion and attitude change (Chaiken, Liberman, & Eagly, 1989; Petty & Cacioppo, 1986). Dual-process approaches distinguish two categories of mechanisms. Mechanisms belonging to the first category (known as ‘system 1’, ‘heuristic’, or ‘associative’) are said to be fast, frugal, unconscious and generally efficient, but prone to mistakes in non-standard circumstances. Mechanisms belonging to the second category (known as ‘system 2’, ‘analytic’, or ‘rule based’) are said to be slow, effort-demanding, conscious and capable of overcoming some of the shortcomings of system 1 processes.

Several dual-process theories have been introduced in the field of reasoning (Evans, 2007; Sloman, 1996; Stanovich, 2004—see Wason & Evans, 1975, for an early suggestion) and in that of decision making (Kahneman, 2003; Kahneman & Frederick, 2002, 2005). Despite their differences (see Osman, 2004), all these theories share a commitment to a broad division of cognitive mechanisms into two categories that helps explain both the shortcomings of ordinary human reasoning, and the fact that these shortcomings can be overcome. Faced, for instance, with a standard reasoning problem (e.g. the famous Wason selection task (Wason, 1966—see below, section 3.2), participants, it is claimed, spontaneously activate their system 1 heuristics that may provide them with a logically inadequate answer (as is common with such abstract and artificial tasks). When they then engage in conscious system 2 reasoning, as they are encouraged to do by the experimental situation, most participants are likely to do so merely to rationalize the solution they already have in mind. Participants who do better are those who are able to discard their initial system 1 inference and to use system 2 mechanisms to solve the problem anew.

The distinction between system 1 and system 2 is fuzzily characterized by contrasting clusters of attributes. It helps explain experimental findings but it lacks theoretical clarity and depth. As stated by Evans, “it would then be helpful to have some clear basis for this distinction” (Evans, 2008, p.270).

Drawing both on advances made possible by dual-process approaches and by evolutionary approaches, we have proposed a principled way of distinguishing two types of inference: intuitive and reflective inferences, without assuming however that they are the output of just two systems (Mercier, submitted; Mercier & Sperber, 2009; Sperber, 1997, 2001; Sperber & Mercier, In press). In this article, focused on the function of reasoning, we just outline our view of the mechanisms involved.

In general, what inferential processes do is produce a representational output warranted by their representational input. When there is an awareness of the output of the inference and of reasons to accept that output, we are dealing with ‘reflective inference’, which correspond to reasoning proper. Most inferences however involve no such awareness and are wholly intuitive. The individual, in whose brain these inferential processes take place at a “sub-personal” level (in the sense of Dennett, 1969), may not even be aware of their output. What an evolutionary approach suggests is that these intuitive inferences, rather than being based on a single inferential mechanism or of constituting a single integrated ‘system’, may well be performed by many domain-specific devices each attuned to the specific demands and affordances of its domain.

The existence and character of many intuitive, domain-specific inferential processes has been particularly well brought to light by work in developmental psychology, for example on “naive mathematics” (e.g., Dehaene, 1999; Wynn, 1992). A 5-month-old infant having seen two dolls placed behind a screen and seeing one doll being taken away is surprised if, when the screen is removed, there are still two dolls present or if no dolls are present at all. Her surprise suggests that she expected there to be only one doll (Wynn, 1992). Where does this expectation come from? It is implausible to think that it is conscious, let alone consciously arrived at, so it must be the output of an intuitive inference. Moreover, the infant is unlikely to have acquired through experience the background knowledge about objects and numbers that would be necessary to draw this inference by means of a domain-general inferential capacity. The hypothesis favoured is that the infant's expectation is derived by a domain-specific ability to draw inferences about the cardinality of small sets of discrete and persistent objects. Similar arguments apply to infants inferential abilities in other domains such as naive physics (e.g., Leslie & Keeble, 1987 ; Spelke & Kinzler, 2007) or naive psychology (e.g., Buttelmann, Carpenter, & Tomasello, 2009; Csibra, Gergely, Bíró, Koós, & Brockbank, 1999; Leslie, 1994; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007).

Work with infants strongly suggests that humans come equipped with several domain-specific inferential mechanisms. Work on adult expertises, for instance chess (Chase & Simon, 1973), shows that other specialized inferential mechanisms are the outcome of an extended learning process. It is plausible that much if not all of human intuitive inference is carried out by a variety of specialized inferential mechanisms, some based more on evolved predispositions and others more on learning, with infant innate mechanisms being at one end, and cultural expertises being at the other end of what may be a continuum of cases.

Humans have the ability to draw inferences not only about ordinary object and events in the world but also about representations of such objects or events. They have, that is, the 'metarepresentational' ability to represent representations, and to draw inferences about them (Sperber, 2000b). Some of these inferences are about mental representations and are essential to our understanding of others and of ourselves (Leslie, 1987). Others are about verbal representations and are essential to our ability to understand meaning in context (Sperber & Wilson, 2002). Other intuitive metarepresentational inferences still are about logical or evidential relationships among representations considered in the abstract.

To use a famous illustration, in Descartes' Cogito, "I think therefore I am" it is intuitively evident that the first proposition, "I think," indeed does entail the second, "I am." What is not at all obvious are the reasons for accepting that this entailment obtains, and philosophers have been hotly debating the issue (e.g., Katz, 1986). The judgment that such an entailment relationship obtains is the output of an intuitive inference. When, on the basis of that intuition, one accepts the conclusion, "I am," one engages in reflective inference, that is, one accepts a conclusion with an awareness of one's reasons for doing so. Such an acceptance is experienced at the personal level as a conscious mental act. In the case of the Cogito, this is true both for Descartes and for his readers. All of us had intuitively taken our own existence for granted. Descartes took a reflective stance to this intuitive belief and found a reason to accept it reflectively, which he then expressed as possibly the most famous argument in the history of philosophy. Likewise, his readers accepting Descartes argument now come to accept its conclusion reflectively (on intuitive and reflective beliefs and on the possibility of believing the same content in both modes, see Sperber, 1997).

In more complex cases of reflective inference, several levels of reasons may be involved. One may for instance accept that the rate of inflation will go up for the reason that the one's favourite newspaper says so, and accept that the its saying so is a good reason to believe so for the further reason that it has been reliable on such issues in the past. But simple or complex, all reflective inferences must ultimately be grounded in an *intuitive* judgment that a given conclusion follows from given premises. In other words, we are suggesting that reflective inference is not a mental mechanism or system that would stand apart from, and in symmetrical contrast to a mechanism or a system of intuitive inference. It is rather an outcome of a specific mechanism of intuitive inference that delivers intuitions about premises-conclusion relationships. When a conclusion is accepted because of such an intuition, this is a mental act of reflective inference or reasoning. When the reflective acceptance of a conclusion itself involves intermediate reflective steps, this makes the whole process a prototypical example of what is commonly meant by reasoning.

Why should the reflective exploitation of one mechanism of intuitive inference among many such mechanisms stand out as so important that it has been seen as what distinguishes humans from beasts? Why should it, in dual-process theories of reasoning, be contrasted on its own with all the mechanisms of intuitive inference taken together? We see three complementary explanations for the saliency of reasoning. First, reasoning is conscious. After all, when we reason, we know that we reason, whereas the very existence of intuitive inference has been a controversial philosophical claim before becoming a cognitive science discovery. Second, while an inferential mechanism that delivers metarepresentational intuitions about premises-conclusion relationship is, strictly speaking, highly domain-specific, the representations about which it delivers intuitions can be representations of anything at all. Therefore, 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 premises-conclusion relationships are domain-specific (as evolutionary psychologists would expect), they have domain general consequences and they 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

It is generally agreed that conscious reasoning is slow, difficult, and not very reliable. This raises a puzzle: what benefit might offset the cost of reasoning and explain its having evolved? Philip Johnson-Laird, who, twenty-five years ago, wrote that “no one knows how deductive competence could have evolved according to the principles of neo-Darwinism” (Johnson-Laird, 1983, p. 142), opines that “no doubt the mind has been shaped by human evolution. But it is hard to tell in what ways” (Johnson-Laird, 2006, p. 256). He agrees however that “evolutionary psychology is a useful heuristic for generating hypotheses” (ibid.) and this is how we intend to use it here. Drawing on evolutionary considerations, we propose a solution to the puzzle of reasoning that not only dovetails with the account we have just outlined of intuitive and reflective inferences but that also should be congenial to all dual-process approaches.

There have been a few cursory attempts in dual-process approaches to explain the function and evolution of reasoning. The main function of system 2 reasoning, it is sometimes claimed (Kahneman, 2003), is to correct for mistakes of system 1 intuition. However, reasoning itself is a potential source of new mistakes. Moreover, much evidence shows 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). However, giving an organism the possibility to deal with novelty and to anticipate the future characterizes not so much reasoning as it does learning (or even, it could be argued, cognition in general). Learning, after all, 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 mistake, or better adapt us to novel circumstances. No doubt, it can. The issue is whether these occasional benefits are such as to explain the costs incurred and hence the very existence of reasoning among humans. It would be unfair and unproductive to push the discussion too far since, manifestly, evolutionary and functional considerations have not been a main concern of dual system theorists (see Mercier & Sperber, 2009). Evolutionary hypotheses are anyhow of little help unless they are precise enough to yield testable predictions and explanations.

Here we want to defend and explore the idea that the function of reasoning is <u>argumentative</u> (Sperber, 2000a, 2001, see also Billig, 1996; Dessalles, 2007; Kuhn, 1992; Perelman & Olbrechts-Tyteca, 1969; Haidt, 2001, offers a very similar take on moral reasoning). We start from a well known problem in the evolution of communication (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984). 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. But, often enough, senders may gain by manipulating receivers and inflicting a cost on them, thus threatening the stability of communication.

In some cases the problem can be solved by the use of signals that provide conclusive evidence of the very information they convey. Signalling one's health or strength by incurring a cost that could not be met by an unhealthy or weak organism is an example where a costly signal is, in itself, an honest indicator of the sender's ability to pay such costs (Zahavi & Zahavi, 1997). However, for most of the uniquely rich and varied information that humans communicate among themselves, there are no available signals, costly or otherwise, that would be proof of their own honesty. Most of human communication consists of cheap talk (for economists, Farrell & Rabin, 1996) or low-cost signaling (for biologists, Maynard-Smith, 1994) the honesty of which is not self-evident. To avoid being victims of misinformation, receiver must exert what may be called "epistemic vigilance" (Mascaro & Sperber, 2009).

Epistemic vigilance can be based on an evaluation of the source of information: People 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 who, out of two possible informants, trust more the one whose expressed views have been more consistent with their own (see Clément, In Press; Harris, 2007, for reviews). The ability to distrust malevolent informants has been shown to develop in stages between the ages of three and six (Mascaro & Sperber, 2009).

Epistemic vigilance can also be based on an evaluation of the communicated information itself through a check of its coherence, and on this we focus. As we use the term, a set of proposition is coherent to the extent that it is cognitively sound to accept all of them together.[ii] Logical inconsistency is a sufficient condition for incoherence, but it is not a necessary one. Just as people accept conclusions on the basis of evidence and background knowledge that fall short of entailing these conclusions, they commonly reject information on weaker grounds than logically inconsistency. Most acceptance or rejection of information is based on non-demonstrative inference.

Effective filtering of communicated information cannot simply consist in rejecting information that is incoherent with one's previous beliefs and goals (even if there is such a tendency is illustrated by the fact that subliminal influence is mostly powerless unless it is coherent with our current goals

(Strahan, Spencer, & Zanna, 2002). This would lead to rejecting valuable updates and corrections, and also to accepting new information that, while being coherent with previously held beliefs, is intrinsically incoherent. A better use of coherence checking consists in accepting or rejecting communicated information in a manner that tends to maximize the coherence of one's resulting beliefs and goals. When a trusted source presents information that is incoherent with previous beliefs, some belief revision is anyhow unavoidable in order to preserve coherence. Either one accepts the new information and should revise previous beliefs on the same topic, or else one should revise one's previous beliefs in the trustworthiness of the source of the information. This may lead to accepting new information that contradicts or otherwise fails to cohere with previously held beliefs and goals (that now have to be revised), provided that the resulting ensemble of mental representation is a more coherent one.

How is coherence checking achieved? Our suggestion is that there is a dedicated metarepresentational mechanism that delivers intuitions about the coherence of sets of representations. In particular it may deliver the intuition that adding a new piece of information to already held background assumptions would result in a loss of coherence, or, on the contrary, that greater incoherence would result from rejecting it. Note that even a modicum of coherence checking involving just newly communicated information and the context of activated background assumption in which it is processed (Sperber & Wilson, 1995) is enough to achieve some serious filtering. We are not suggesting that coherence checking is applied on a wider scale, or to the individual's own background knowledge in the absence of a challenge presented by communication.

Assume that humans, as receivers of information, are not simply gullible and, often enough, exert their epistemic vigilance by checking coherence as we have just outlined. What then are the options of a communicator addressing such a vigilant audience? She might want to communicate a piece of information that the addressee is unlikely to accept for it will seem to him more coherent to doubt her reliability than to accept what she says. One way to overcome this difficulty would be to provide evidence of her reliability in the matter (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 way to convince her addressee is to address directly and overtly his coherence concern by showing that it would less coherent for him to reject her claims than to accept them. How can this be done? By arguing, and more specifically, by presenting premises that the addressee already believes or is willing to accept on trust and by showing that, once these premises are accepted, it would be less coherent to reject the conclusion than to accept it. Doing so is, of course, engaging in explicit reasoning and encouraging the audience to follow this reasoning. Such explicit reasoning, or arguing, is a way to try and convince an epistemically vigilant audience.

The massive dependence of humans on information obtained from others through verbal communication creates a selective pressure for epistemic vigilance. Vigilance can be, and is exerted both towards the source and towards the content of communicated information. Vigilance towards the content takes the form of coherence checking. Having to address coherence checkers creates a pressure on communicators to provide information that can withstand such a check. It also creates an opportunity to highlight the contribution of the new information to the coherence of the audience's views, that is, to argue for it. This in turn creates a pressure on communicators to develop the reasoning skills needed for argumentation. Being presented with arguments creates a pressure on the audience to develop the reasoning skills needed to evaluate arguments. Reasoning, in this perspective evolves as means to convince interlocutors through arguments and, for the audience, to accept good arguments and reject bad ones (Mercier, submitted; Mercier & Sperber, 2009). People being communicators and addressees in turn are under convergent pressures to reason.
In communication, and more precisely in the production and evaluation of verbal arguments, reasoning fulfils a function the importance of which is in proportion to the importance of communication itself. This function cannot be fulfilled with a comparable efficiency in any other way available to humans. Of course, as we pointed out, epistemic vigilance is exerted not only towards the content but also towards the source of information, and communicators can convince their audience through rhetorical means of seduction or intimidation. But the fact is that argument can achieve what no other rhetoric device can achieve. People whom we would never believe on trust on a given issue may still convince us by means of a compelling argument. Both the benefits and the risks involved in our dependence on communication are huge. By increasing these benefits and lowering these risks, even if to a limited degree, the social practice of argumentation and its cognitive tool, reasoning, well compensate their cost.

Thus according to the present theory the function of reasoning is <u>argumentative</u>: to find and evaluate reasons so as to convince others and be convinced only when appropriate. This hypothesis grounds the evolution of reasoning in that of communication. It makes reasoning a cognitive mechanism aimed at social interaction, in line with much current work that has stressed the role of sociality in human unique cognitive capacities (Byrne & Whiten, 1988; Dunbar, 1996; Dunbar & Shultz, 2003; Hrdy, 2009; Humphrey, 1976; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Whiten & Byrne, 1997). To say that the function of reasoning is argumentative implies that it is tailored for this function; it does not imply that it cannot be also used otherwise, in particular in individual ratiocination. Still, even there, much reasoning is done as if addressing an imaginary or a future audience (as suggested, for instance, by Piaget, 1928).

This view has implications for the philosophy of mind (Mercier & Sperber, 2009; Sperber & Mercier, In press) and for the analysis of reasoning mechanisms (Mercier, submitted). It entails testable predictions regarding the way reasoning should work, the effects it should have, and the contexts in which it should be triggered. Our goal here is to spell out some of these predictions, to evaluate them in the light of the available evidence and to show how they help make better sense of a number of puzzles in the psychology of reasoning and of decision making (many of which are already better understood thanks to dual-process approaches).

The most straightforward prediction of the argumentative theory is that people should be reasoning at their best when this serves the function of producing or of evaluating arguments in a proper argumentative context. In the next section, evidence will be presented that humans are indeed skilled arguers. In section three, we will turn to the confirmation bias. This bias, a puzzle for classical theories of reasoning, has a straightforward explanation within the present framework: it contributes to the effectiveness of argumentation. The study of motivated reasoning will be used, in section four, to show how reasoning often buys social advantages at an epistemic price. Finally we will consider a question suggested by the argumentative approach: does reasoning allow people to make better decisions, or decisions that are easier to justify?

2 Argumentative skills

2.1 Understanding and evaluating arguments

The understanding of arguments has been studied in two main fields of psychology: persuasion and attitude change, and reasoning. The aims, methods and results are different in these two fields. Operating 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 arguments (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 confronted 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 other participants 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 those participants who would be directly affected by the setting up of a comprehensive exam were much more influenced by strong than by weak arguments.

These experiments have shown that many factors influence the effectiveness of arguments, some of which seem to be unrelated to reasoning, such as the gender of the source for example. Dual process theories account for these findings by postulating that arguments can affect their target through two types of mechanisms (Chaiken et al., 1989; Petty & Cacioppo, 1986). A central route—reasoning—allows people to evaluate the strength of the arguments while simpler cues—such as the attractiveness of the speaker—affect persuasion through more peripheral routes. Two overall findings, illustrated in the experiment above, are relevant here. First participants are able to use reasoning to accurately evaluate arguments: they are more influenced by strong than by weak arguments (see (Petty & Wegener, 1998) for a review). Second, their doing so is conditional on their motivation, and not any motivation will do: "Perhaps the most important variable influencing a person's motivation to think is the perceived personal relevance or importance of the communication" (Petty & Wegener, 1998, p. 328, see B. T. Johnson & Eagly, 1989; Petty & Cacioppo, 1979, 1990). This should not be construed as a flaw: given that reasoning is so costly (Gailliot et al., 2007; Masicampo & Baumeister, 2008), it is only sensible that its use should be conditional on the relevance of the argument at hand.

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 confronted with premises and asked to either produce or 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". The non-logical content of these problems is generally as limited as possible, precisely in order to avoid interference from general knowledge. For such tasks, Evans recognizes that "logical performance [...] is generally quite poor" (Evans, 2002, p. 981). To give but one example, it was found in a review that an average of 40% of participants fail to make the simple modus tollens inference that was taken as example (if p then q, not q, therefore not p) (Evans et al., 1993). This level of performance can be easily explained by taking motivation into account. Given that the content of reasoning problems is almost never of any personal significance, the research on persuasion predicts that people should not reason while evaluating reasoning problems (!) but should instead rely on heuristics.[iii] As soon as the content is made somewhat relevant, performances improve dramatically. For instance, people can understand modus tollens arguments as a matter of course in a more natural context (see Scholnick & Wing, 1991; V. A. Thompson, Evans, & Handley, 2005). These results strengthen the conclusion that a natural, felicitous argumentative context is conducive to a proper activation of reasoning.

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 the context in which they are used they can be more or less fallacious. Here is an example with a variation on the argument from ignorance (ad ignorantiam). In an experiment, participants had to rate the strength of the following argument:

> Barbara: Are you taking digesterole for it?
> Adam: Yes, why?
> Barbara: Good, because I strongly believe that it does not have side effects.
> Adam: It does not have any side effects.
> Barbara: How do you know?
> Adam: Because I know of an experiment where they failed to find any.
> (Oaksford & Hahn, 2004, p. 80)

The fallaciousness of the ad ignorantiam argument in this case is an inverse function of the number of experiments that failed to find side effects. Accordingly, participants rated the above argument as weaker than a variant in which Adam knew of 50 experiments that had failed to find side-effects (see also (Hahn, Oaksford, & Bayindir, 2005).

Various experiments have shown that participants are generally able to spot other argumentative fallacies (Hahn & Oaksford, 2007, experiment 3, Neuman, 2003; Neuman, Weinstock, & Glasner, 2006; Weinstock, Neuman, & Tabak, 2004). Not only do they spot them, but they tend to react appropriately: rejecting them when they are indeed fallacious (which they are not always) or being convinced to the degree that they are well grounded. This is the case for the slippery slope (Corner, Hahn, & Oakfsord, 2006) and circular reasoning (Hahn & Oaksford, 2007; Rips, 2002). It has even been shown that children as young as six favour non-circular arguments (Baum, Danovitch, & Keil, 2007). When researchers have studied other skills specific to argumentation, performances have 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 appropriately attribute the burden of proof (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.[iv] 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 having been 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 flaw is that people resort to mere explanations (‘make sense’ causal theories) instead of relying on genuine evidence (data) in support of their views. However, later research has shown that this is mostly an artefact of the lack of evidence available to the participants: when evidence is made available, participants will favour it (both in production and evaluation) (Brem & Rips, 2000, see also Hagler & Brem, 2008; Sá, Kelley, Ho, & Stanovich, 2005). A second flaw noted by Perkins and Kuhn is the relative superficiality of the arguments used by the 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 might have been. 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 afterwards a significant improvement in the quality of the arguments they used (Kuhn, Shaw, & Felton, 1997, see Blanchette & Dunbar, 2001, for similar results regarding analogical reasoning).

The third flaw, noted by Perkins and Kuhn, is the most relevant one here. Participants had trouble generating counter-arguments against their own claims. Seen from an argumentative perspective, this may be however, rather than a simple flaw, a feature of argumentation that contributes to its effectiveness in fulfilling its function. If one’s goal is to convince others, one should first and foremost look for supportive arguments. Looking for counter-arguments against one’s own claims may be part of a more sophisticated and effortful argumentative strategy aimed at anticipating the interlocutor's response, but, in the experimental setting, there was no back-and-forth to encourage such an extra effort. If this is a correct explanation of what need not be a flaw after all, then the difficulty people seem to have in coming up with counter-arguments should be easily overcome by having them challenge someone else's claims rather than defend 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 counter-arguments against it (Kuhn, Weinstock, & Flaton, 1994). In another experiment all participants were able to find counter-arguments against a claim (that was not theirs), and to do so very quickly (Shaw, 1996).

When people have looked at reasoning performances in felicitous argumentative settings, they have observed good performances. Resnick and her colleagues created groups of three participants who disagreed on a given issue (Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993). Analyzing the debates, the researchers were "impressed by the coherence of the reasoning displayed. Participants […] appear to build complex argument 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. 362-3, see also Blum-Kulka, Blondheim, & Hacohen, 2002; Hagler & Brem, 2008; Stein, Bernas, & Calicchia, 1997; Stein, Bernas, Calicchia, & Wright, 1995). By looking at the details of the arguments used, it is possible to draw a comparison with participants' poor performance in reasoning experiments. The modus tollens for instance, that confuses nearly half of the participants in decontextualized tasks, is "surprisingly common" (and appropriately used) in argumentative contexts (Pennington & Hastie, 1993, p.155).

A similar pattern emerges from developmental studies. Nancy Stein and her colleagues have shown, through analysis of corpora and experimental studies, that children as young as three engaged in a discussion are able "to generate and think about positive and negative reasons for pursuing different courses of action or for holding specific sets of beliefs" (Stein & Bernas, 1999, p. 97, see also Clark & Delia, 1976; Stein & Miller, 1993). Dunn and Munn (1987) even demonstrates that 2 year-olds produce appropriate justifications (see (Mercier, In prep) for further evidence regarding young children). As with adults, these good performances were obtained because the researchers focused on situations that are "personally meaningful to young children and [...] that impact directly on their goals, beliefs, and well-being" (Stein & Bernas, 1999, p. 97). Lastly, children display the same biases as adults: "Arguers of all age levels, from preschool to adulthood, [...] exhibit similar biases in their understanding and memory for a conflict, independent of their age." (Stein & Albro, 2001, p.130).

Given that most standard reasoning tasks could not even be used with children this age, these results are quite striking. They strengthen the idea that reasoning is very natural in argumentative context, and only in argumentative contexts.

2.3 Group reasoning

If people are skilled both at producing and at evaluating arguments, and if these skills are displayed most easily in argumentative settings, then debates should be especially conducive to good reasoning performances. Many types of tasks have been studied in group settings, yielding very mixed results (see Kerr, Maccoun, & Kramer, 1996; Kerr & Tindale, 2004, for recent reviews). The most relevant findings here are those pertaining to logical or, more generally, intellective 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 (pre-test) then solve the same problems in groups of 4 or 5 members (test) and then solve them individually again (post-test), to make sure that any improvement does not come from mere following of other group members. Their performances are compared to those of a control group of participants who underwent the same tests, but always individually. Intellective 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 (B. L. Bonner, Baumann, & Dalal, 2002; Laughlin & Ellis, 1986; Stasson, Kameda, Parks, Zimmerman, & Davis, 1991)[v]. This can lead to large improvements in performances. Some experiments that used the Wason selection task dramatically illustrate this phenomenon (Moshman & Geil, 1998, see also Maciejovsky & Budescu, 2007). The Wason selection task is the most widely used task in reasoning and the performances of participants are generally so low as to be barely distinguishable from the level expected at chance of 6.25% 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.[vi] This is the normal outcome of the conjunction of two factors: the ability of the participants who have understood the problem (or parts of it) to produce convincing arguments, and the ability of the other participants to be selectively swayed by these arguments.

Several challenges can be levelled against this interpretation of the data. It could be surmised that the person who has the correct solution simply points it out to the others who immediately accept it without arguments, perhaps because they have recognized this person as the 'smartest' (Oaksford, Chater, & Grainger, 1999). A simple look at the transcripts of the experiments shows this not to be the case (see for instance Moshman & Geil, 1998; Trognon, 1993). More generally, many experiments have shown that debates are essential for any improvement of the performances in group settings (see Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006, for a review and some new data, and Mercier, In prep, for similar evidence in the development and education literature). Moreover, in these contexts, participants decide that someone is smart based on the strength and relevance of her arguments, not the other way around (Littlepage & Mueller, 1997). 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 to 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; Laughlin et al., 2006; Laughlin et al., 2003; Lombardelli, Proudman, & Talbot, 2005; Michaelsen, Watson, & Black, 1989; Sniezek & Henry, 1989; Stasson et al., 1991; Tindale & Sheffey, 2002), a phenomenon also known as "two wrongs make a right" in the developmental literature (Ames & Murray, 1982; Glachan & Light, 1982; Schwarz, Neuman, & Biezuner, 2000).

Another counter-argument claims that people are simply more motivated, generally, when they are in groups (Oaksford et al., 1999). This is not so. 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). Still, even if group settings are not conducive to an overall increase in performance for a large range of tasks, one might argue that group settings specifically motivate people to perform well on intellective tasks. If this is the case, there are two possible explanations. The increase in the motivation to reason could be specific to argumentative setting and processes—which is the hypothesis defended here. Or it could be a more general motivational phenomenon: higher motivation generally enhances performance and this should be true also in the area of reasoning. It is easy to disentangle these two hypotheses by looking at whether other incentives that could reinforce motivation have similar effects on reasoning. The closest comparison is provided by an experiment in which participants were offered monetary rewards for a correct answer for the Wason selection task (Johnson-Laird & Byrne, 2002, see also Jones & Sugden, 2001). No improvement was observed. By and large monetary incentives have no effect on the performances in reasoning and decision making tasks (Ariely, Gneezy, Loewenstein, & Mazar, In Press; S. E. Bonner, Hastie, Sprinkle, & Young, 2000; S. E. Bonner & Sprinkle, 2002; Camerer

& Hogarth, 1999). This means that not any incentive will do: group settings have a motivational power to which reasoning responds specifically.

The argumentative theory also helps predict what will happen in non-optimal group settings. If all group members share an opinion, a debate should not spontaneously occur. However, many experimental and institutional settings (juries, committees) force people to discuss, even if they already agree. When all group members agree on a view, each can find arguments supporting this view. These arguments will not be critically examined, let alone refuted, thus offering other group members additional reasons to hold their views. This should lead to a strengthening of the opinions (see Sunstein, 2002, for a review, and Hinsz, Tindale, & Nagao, 2008, for a recent illustration). Contra Sunstein's 'law of group polarization', one has to bear in mind that this 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 depolarisation (Kogan & Wallach, 1966; Vinokur & Burnstein, 1978). In both cases, the behaviour 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. These good performances offer a striking contrast with the poor results obtained in abstract reasoning tasks. Finally, the increase in performance observed in argumentative settings confirms that reasoning is at its best in these contexts. We will presently 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 and evidence of its prevalence and robustness abound (see Nickerson, 1998, for review). 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 and hence a feature of reasoning when used for the production of arguments.

In fact, we suggest, the label of 'confirmation bias' refers to two distinct types of cases, both characterized by the failure to look for counter-evidence or counter-arguments to an existing beliefs, both predicted by the argumentative approach, but brought about in different ways.
In one type of cases, we are dealing not with flawed reasoning but with absence of reasoning. Such absence of reasoning is to be expected when people already hold some belief on intuitive ground and do not have to argue for it. In such cases, people may express as predictions expectations that intuitively follow from their intuitive beliefs. Ordinarily, these are positive predictions, that are both easier to derive and more relevant. For instance, if I believe that I have left my keys in my trousers, I am likely to spontaneously make the easy and relevant inference that this is where they still are. Except in special circumstances (for instance if it had been suggested that my keys were in my jacket), it would take some more thinking and be less relevant to infer from my belief that my keys are not in my jacket, even though both inductive inferences are equally valid. 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).

In another type of cases, people are in a situation to argue for some claim of theirs. In that case, the relevant arguments are those that support their claims or undermine the counterarguments of their interlocutors. People arguing will actively look for evidence confirming their own claim, but, while this may be seen as a bias from a normative epistemological point of view, it clearly serves the goal of convincing others.

It is possible to make three broad predictions. The first is that the true confirmation bias (as opposed to plain trust in one's intuitive beliefs and their positive consequences) should occur only in argumentative situations. The second prediction is that it should be limited to the production of argument. The rationale for the existence of a confirmation bias in the production of arguments in favour of a given claim does not extend to the evaluation of arguments by an audience that is just aiming at being well informed. The third prediction is that the confirmation bias in the production of arguments is not a bias in favour of confirmation in general and against disconfirmation in general, but is a bias in favour of confirming one's own claims, a bias that should be naturally complemented with a bias in favour of disconfirming opposite 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 regarding the rule and test it by proposing a triple of their own choice. The experimenter says whether this triple conforms to the rule or not. Participants can repeat this until they feel ready to venture their hypothesis about the rule. The experimenter tells them whether or not their hypothesis is true. If it is not, participants 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, it is, we suggest, the intuitive way to exploit one's intuitive hypotheses, checking for instance that our keys are where we believe we left them rather than checking that they are not where it follows from our belief that they should not be. What we see here is then a sound heuristic rather than bias.

Still, participants in the (2, 4, 6) task are acting ineffectively, wasting time on mistaken hypotheses that it would have been easy to disconfirm. This result, we suggest, from the fact that the task is contrived so that it elicits mistaken intuitions. The trick of the experiment is that the actual rule, "ascending numbers", is much more general than the rule made intuitively more plausible by the initial triple. Why should a hypothesis more specific than the correct one be so intuitive? Because the initial example exhibits striking regularities (even numbers, equal interval between the numbers, ascending order) all of which seem prima facie relevant when, in fact, only the last one of these regularities is. Intuitive hypothesis formation is guided by apparent relevance towards hypotheses that are, in this case, too specific. In this rare situation, the usual "positive test" heuristic turns out to be ineffective. Still, its general effectiveness argues against describing it as a "bias", and since it involves no reasoning, it is no reasoning bias anyhow.

What is striking is the failure of attempts to cause participants to reason in order to correct their ineffective approach. It has been shown that even when instructed try to falsify the hypotheses they generate, less than one participant in 10 is able to do so (Poletiek, 1996; Tweney et al., 1980). Given that 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. If, however, a hypothesis is presented as coming from someone else, more participants, it seems,[vii] will try to falsify it and they will give it up in favour of another hypothesis much more readily (Cowley & Byrne, 2005). The same applies if the hypothesis is that of a minority member in a group setting (Butera, Legrenzi, Mugny, & Pérez, 1992). So falsification is accessible provided that the situation encourages arguing against an hypothesis other than one's own.

3.2 The Wason selection task

A similar interpretation can be offered to account for results obtained with the Wason selection task (Wason, 1966). In this task, participants are given a rule regarding four cards. In the original version, the cards have a number on one side and a letter on the other, with only one side being visible—they might read, 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 tell the cards that need to be turned over in order to determine if the rule is true. In this task too it is useful to distinguish the effect of intuitive mechanisms from that of reasoning proper (as long suggested by Wason and Evans (1975). Intuitive mechanisms involved in understanding utterances will draw the participants' attention towards the cards that are made most relevant by the rule and the context (Girotto, Kemmelmeier, Sperber, & Van der Henst, 2001; Sperber, Cara, & Girotto, 1995). In the standard case, these will simply be the cards mentioned in the rule (the vowel, E, and the even number, 4) instead of the correct answer (the E and the 7). Given that the E and the 4 can only confirm but not falsify the rule the behaviour of participants choosing selecting cards could be interpreted as an example of 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 do not aim either at 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 participant's attention has been directed towards some of the cards and they have arrived at an intuitive answer to the question, reasoning occurs not to evaluate and correct their initial intuition, but to find justifications for it (Evans, 1996; Lucas & Ball, 2005; Roberts & Newton, 2002). 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 performances in group settings mentioned earlier. Participants should also be able to try to falsify the rule when their first intuition is to think the rule is false and when they want to prove it wrong. Researchers have used rules such as "all members of group A are Y", with Y being a negative or a positive stereotype (Dawson, Gilovich, & Regan, 2002). Participants most motivated to prove the rule wrong—those belonging to group A when Y was negative—were able to reach 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". Despite being solvable by very simple programs (see for instance 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, Handley, Harper, & Johnson-Laird, 1999). This constitutes the participants' initial intuition. In order 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. 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, Musch, & Naumer, 2000; Newstead, Handley, & Buck, 1999).

Again, we suggest, this should not be interpreted as lack of ability, but of motivation. Paradoxically, one case in which participants look for alternative models is when they think there are none. In some experiments, participants have to decide on the possibility of a given conclusion. Sometimes, their initial intuition is that this conclusion is impossible. In order to prove that they are right, they search for possible ways to arrive at this conclusion, expecting there will be none and thus making their point (Evans et al., 1999). This illustrates the fact that when a search for other possibilities does serve participants' argumentative goals, then they are able to carry it out.

Participants have other means to falsify a conclusion when they want to prove it wrong. This occurs for normal conclusions presented by someone else (Sacco & Bucciarelli, 2008) or when participants are confronted 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 they actually perform much better when the conclusion is unbelievable (see for instance (Evans, Barston, & Pollard, 1983). It is not that participants reason more in this case—they spend as much time trying to solve problems with believable than with unbelievable conclusions (V. A. Thompson, Striemer, Reikoff, Gunter, & Campbell, 2005). It is that the direction reasoning takes is mostly determined by the initial intuitions of the participants. 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 reviewed—hypothesis testing, the Wason selection task, and syllogistic reasoning—a similar pattern can be observed. Participants have intuitions that lead them towards some answers. Reasoning, if it is used at all, is mostly used to confirm these initial intuitions. Let us stress again that the confirmation bias does not show that people are bad at falsifying: when people want to prove a statement wrong, they are quite able to use falsifying strategies. This is exactly what one should expect of an argumentative skill. However, seeing the confirmation bias as a normal feature of reasoning playing a role in the production of arguments can be surprising in light of the poor outcomes it has been deemed to cause. Conservatism in science is one example (see (Nickerson, 1998) and references within). Another is the related phenomenon of groupthink that has been held responsible for many disasters, from the Bay of Pigs fiasco (Janis, 1982), to the tragedy of the Challenger shuttle (Esser & Lindoerfer, 1989; Moorhead, Ference, & Neck, 1991) (see Esser, 1998, for review). In such cases, reasoning tends not to be used in its normal context, i.e. 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. When reasoning is used in a more felicitous context however, that is, in arguing among people who disagree but have a common interest in the truth, the confirmation bias contributes to an efficient form of division of cognitive labour.

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 expound these arguments to the group, to be tested by the other members. Such a method will work as long as people are able to be swayed by good arguments, and the results reviewed in section 2 show this to be generally the case. This joint and dialogic approach is much more efficient than one where each individual on his or her own has to examine carefully all possible solutions[viii]. The advantages of the confirmation bias are even more salient when each participant in a discussion is in a better position to look for arguments in favour of his or her favoured solution. This will typically happen when people want to solve a collective action problem. A wife and her husband have to decide what to do on a given evening. Each can find arguments for her or his favoured solution, arguments that may tap into unshared information: "I've seen that movie, it would be nice to watch something else", "I haven't had Japanese for a while, we should go there", etc. Given that it would be much more costly, even impossible in many cases, for each individual to find arguments for the other side, it is then beneficial to focus on one's side—to have a confirmation bias. Again, as long as each speaker can state her or his position freely and the other is sensibly influenced by her or his arguments, this will lead to a good outcome at a low cognitive price.

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 one another, making it hard to properly assess the two processes independently of one another. 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 it were not so, 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 a communication context, when the addressee trusts the communicator enough to pay attention, but not enough to immediately accept what she claims. In such a case, the addressee is likely to be vigilant towards the content communicated and to check its coherence, and the speaker is likely to argue for claims she seriously wants the addressee to accept. People can also be proactive and anticipate situations in which they might have to argue to convince others that their claims are true or, in other situations of relevance here, that their actions are justified. Much reasoning, we would say, anticipates the need to argue. We will show in this section that work done on motivated reasoning can usefully be reinterpreted in this perspective, and, in the next section, that the same holds for work done on reason-based choice.

Many of our beliefs are likely to remain unchallenged because they are of relevance only to us 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 our beliefs are unlikely to be universally shared or to be accepted on trust just because we express them. When we pay attention to the contentious character 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 in a position to state them. If the search for arguments is successful, we will be ready. If it is not, then perhaps it might be better to adopt a weaker position, easier to defend. Such uses of reasoning have been intensively studied under the name of motivated reasoning[ix] (Kunda, 1990, see also Kruglanski & Freund, 1983; Pyszczynski & Greenberg, 1987, and Molden & Higgins, 2005, for a recent review).

4.1 Motivated reasoning

A series of experiments by Ditto and his colleagues, involving reasoning in the context of a fake medical result, will illustrate the notion of motivated reasoning (Ditto & Lopez, 1992; Ditto, Munro, Apanovitch, Scepansky, & Lockhart, 2003; Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998). Participants had to put some saliva on a strip of paper and were told that its change of colour, or lack thereof depending on the condition, would be an indication of an unhealthy enzyme deficiency. Participants, being motivated to believe they were healthy, tried to garner arguments for this belief. Those, for instance, who had been told that a change of colour was a good sign waited longer to see whether the strip would change colour, and they often resorted to dampening it again. After having taken the test, participants were asked to rate its efficiency. In one version of the experiment, the rate of false positives was given and varied across conditions. The use that was made of this information reflects motivated reasoning. When the rate of false positives was high, participants motivated to reject the conclusion used it in order to diminish the validity of the test. This same high rate of false positives was discounted by participants 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, thereby providing them with 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. If participants were just disposed to wishful thinking, they might ignore or dismiss the test. What they do instead is look for evidence and arguments showing that they are healthy, or at least for reasons to question the value of the test. It is possible but not self-evident that they are trying to convince themselves. It is clear at least that they are developing arguments with which they could defend their point of view against the challenge of others.

Other studies have demonstrated the use of motivated reasoning in order to support various beliefs that others might challenge. Participants dig in, and occasionally alter their memories to support a positive view of themselves (Dunning, Meyerowitz, & Holzberg, 1989; M. Ross, McFarland, & Fletcher, 1981; Sanitioso, Kunda, & Fong, 1990). They modify their causal theories to defend some favoured belief (Kunda, 1987). After being informed of the outcome of a game on which they had made a bet, they use events of 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 when they don't agree with its conclusions, in order to justify its rejection (Koehler, 1993; Mahoney, 1977). In economic settings, people use information in a flexible manner so as to be able to justify their preferred conclusions, or make the decision they favour (Boiney, Kennedy, & Nye, 1997; Hsee, 1995, 1996a; Schweitzer & Hsee, 2002).

All of these experiments demonstrate that people sometimes look for reasons to justify an opinion they are eager to uphold. From an argumentative point of view, they do so not to convince themselves of the truth of their opinion but to be ready to meet the challenge of others. If they fail to achieve such a readiness, they may become reluctant to express an opinion they are unable to defend, and less favourable to that 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 aiming at epistemic benefits, its use to justify an opinion already held is hard to explain, especially since, as we now show, motivated reasoning can have dire epistemic consequences.

4.2 Consequences of motivated reasoning

4.1.1 Biased evaluation and attitude polarization

In a landmark experiment, Lord and colleagues asked participants who had been previously selected as being either proponents or opponents of the death penalty to evaluate studies bearing on its efficiency as a deterrent (Lord, Ross, & Lepper, 1979). The studies given to the participants had different conclusions: while one was showing the death penalty to have a significant deterrent effect, the other was yielding an 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 has having been much more poorly conducted. In this case, participants used reasoning not so much to objectively assess the studies as to confirm their initial views by alternatively finding flaws and strength in similar studies, depending on their conclusion. This phenomenon is known as biased assimilation or biased evaluation. Note however that participants are not trying to form an opinion: they already have one. Rather, we suggest, they are considering arguments and counterarguments to defend their opinion. Their goal is argumentative rather than epistemic and ends up being pursued at the expense of epistemic soundness.

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 stems from the trivial fact that rejecting what we are told generally requires justifications whereas accepting it does not. Moreover, the time spent on these arguments is mostly dedicated to finding counterarguments (Edwards & Smith, 1996, see also Brock, 1967; Cacioppo & Petty, 1979; Eagly, Kulesa, Brannon, Shaw, & Hutson-Comeaux, 2000). Participants tend to comb arguments for flaws, and do find them, 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, Gordon, & Fauth, 1997; Klaczynski & Lavallee, 2005), or argumentative fallacies (Klaczynski, 1997). In all of these cases, motivated reasoning leads to a biased assessment: arguments with unfavoured conclusions are rated as less sound and as less persuasive than arguments with favoured conclusions.

Sometimes the evaluation of an argument is biased to the point of having an effect opposite to the one intended by the arguer: upon reading an argument with a counter-attitudinal conclusion, interlocutors may find so many flaws and counter-arguments that their initial unfavourable attitude is in fact strengthened. This is the phenomenon of attitude polarization that has been extensively studied following its first demonstration by (Lord et al., 1979, see also Greenwald, 1969). Due to methodological constraints,[x] the phenomenon has proven hard to replicate (Kuhn & Lao, 1996; Miller, Michoskey, Bane, & Dowd, 1993). Pomerantz and colleagues, however, managed to show that individuals holding strong attitudes with great confidence will tend to move towards more extreme attitudes after reading a counter-attitudinal argument (Pomerantz, Chaiken, & Tordesillas, 1995). Even more interestingly, Taber and Lodge have demonstrated that, in the domain of politics, attitude polarization could be observed for participants who not only hold strong attitudes but who also are most knowledgeable (Taber & Lodge, 2006, see also Redlawsk, 2002). Their knowledge allows these participants to find more counter-arguments, leading them to more biased evaluation. It should be stressed that in all of these experiments, although they are supposed to bear on the evaluation of arguments, participants spend more time producing than evaluating arguments. Strictly speaking, the evaluation phase ends once the argument is read. This phase, by itself, should not show a bias. However, if participants disagree with the conclusion of the argument, they will then try to produce counter-arguments. It is in this process of finding these counter-arguments that they will dissect the argument to find flaws in it. So even though this process results in a biased assessment of arguments, this is an effect of biased production of arguments, not biased evaluation as such.

4.1.2 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 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 & Conlee, 1975) showed that polarization increases with the time spent thinking about an item, and Jellison et Mills (1969) that it increases with the motivation to think. As in the case of polarization following biased evaluation, polarization occurs only when participant are knowledgeable. For instance, women’s (but not men’s) attitudes towards fashion related items showed polarization, and the converse was true for sports (Tesser & Leone, 1977, see also Millar & Tesser, 1986). It is possible to mitigate this effect 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, 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 towards which they intuitively have a positive or negative attitude, what happens, we suggest, is that they reflect less on the item itself than on ways to defend their initial attitude. Many other experiments have showed that once people have formed an attitude about a target, they will look for information that supports this attitude (Hart et al., in press; S. M. Smith, Fabrigar, & Norris, 2008) and try to integrate any presented information in its support as well (Bond, Carlson, Meloy, Russo, & Tanner, 2007; Brownstein, 2003), leading them to choose inferior alternatives (Russo, Carlson, & Meloy, 2006).

According to the argumentative theory, reasoning should be even more biased once the reasoner has already stated her opinion, thereby increasing the pressure to justify it instead of moving away from it. This phenomenon is called bolstering (W. J. McGuire, 1964). Thus, when participants are committed to an opinion, thinking about it will lead to a much stronger polarization (Lambert, Cronen, Chasteen, & Lickel, 1996; Millar & Tesser, 1986). Accountability (the need to justify one's decisions) will also increase bolstering (Tetlock, Skitka, & Boettger, 1989), see (Lerner & Tetlock, 1999) for review). Such bolstering can increase some decision errors, such as the sunk cost fallacy (Conlon & Wolf, 1980; Fox & Staw, 1979).

Finally, motivated reasoning should also have effects on confidence. When participants think of an answer for a given question, they will spontaneously be tempted to generate reasons supporting that answer. This may then cause overconfidence in the answer. Koriat and his colleagues (Koriat, Lichtenstein, & Fischhoff, 1980) have tested this hypothesis using general knowledge questions such as "the Sabines were part of (a) ancient India or (b) ancient Rome." After having answered 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. In the first group, there was no difference with the baseline condition: both groups were equally overconfident. This suggests that thinking of reasons supporting their answers is what people spontaneously do anyhow when they think of 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 reasons against their answer allowed them to see its limitations, something they would not do on their own (see Arkes, Guilmette, Faust, & Hart, 1988; Davies, 1992; Griffin & Dunning, 1990; Hirt & Markman, 1995; Hoch, 1985; Yates, Lee, & Shinotsuka, 1992, for replications and extensions to the phenomenon of hindsight bias and the fundamental attribution error). It is then easy to think that overconfidence would also be alleviated by having participants discuss their answers with someone favouring a different conclusion.

These results may be unsurprising but they raise a problem for the classical view of reasoning. In all these cases, reasoning does not lead to more accurate attitudes about an object, or a better estimation of the correctness of one's answer. Instead, by looking only for supporting arguments, reasoning strengthens people's opinions and distorts their estimations. Epistemic goals, in these cases, are not well served by reasoning. On the other hand, argumentative goals are.

4.1.3 Belief perseverance

Motivated reasoning can also be used to hang on to beliefs despite their having been proven to be ill-founded. This phenomenon, known as belief perseverance, is "one of social psychology's most reliable phenomena" (Guenther & Alicke, 2008, p.706). In a classic demonstration, Ross and his colleagues (L. Ross, Lepper, & Hubbard, 1975) gave participants a feedback on their performances. Afterwards they told them that in fact the feedback was completely bogus, an experimental trick designed to study people's reactions to feedback. Nevertheless, participants' self-evaluations following the debriefing were strongly influenced by the discredited feedback. The involvement of motivated reasoning in this effect can be demonstrated by providing participants with evidence both for and against a favoured belief. If belief perseverance were a simple result of some psychological inertia, then the first evidence presented should be the most influential, no matter what its direction might be. On the other hand, if evidence can be used selectively then only evidence supporting the favoured belief should be retained, regardless of the order of presentation. Guenther and Alicke (2008) tested this hypothesis in the following manner. 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. Positive or negative feedback was then given to the participants, but a few minutes later it was explained that the feedback was actually bogus and the real aim of the experiment was explained. At three different points the participants had to evaluate their performances: 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, thereby allowing for the maintenance of a positive view of their performances. Those who had received negative feedback, however, behaved in an opposite manner: they took the debriefing fully into account, thus rejecting the negative feedback and restoring a positive view of themselves. This strongly suggests that belief perseverance of the form just described is an instance of motivated reasoning (see Prasad et al., 2009 and Nyhan & Reifler, In prep. for applications to the domain of political beliefs).

5 Proactive reasoning in decision

In the previous section, we have argued that much reasoning is done in anticipation of situations in which an opinion might have to be defended and we have suggested that work on motivated reasoning can be fruitfully reinterpreted in this light. Even more than their opinions, people may have to put forward arguments to defend their decisions and actions and they may reason proactively to that end. This, we want to argue, is the main role of reasoning in decision making. Such a claim stands in sharp contrast to the classical view that reasoning on possible options and weighting their pros and cons is the most if not the only reliable way to arrive at sound decisions (Janis & Mann, 1977; Kahneman, 2003; Simon, 1955). This classical view has, anyhow, be vigorously challenged in much recent research. Some argue that the best decisions are based on intuition and made in split seconds (see for instance 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, Bos, Nordgren, & van Baaren, 2006; Dijksterhuis & van Olden, 2006). We briefly review these challenges to the classical view before considering the important literature on reason-based choice and interpreting it in the light of the argumentative theory of reasoning.

5.1 To what extend does reasoning help deciding?

In a first series of studies, Wilson and his colleagues looked at the effect of reasoning on the consistency between attitudes and behaviour (Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984; Wilson, Kraft, & Dunn, 1989; Wilson & LaFleur, 1995, see also Koole, Dijksterhuis, & Van Knippenberg, 2001; Millar & Tesser, 1989; Sengupta & Fitzsimons, 2000; Sengupta & Fitzsimons, 2004, and Wilson, Dunn, Kraft, & Lisle, 1989, for review). In one of their first experiments (Wilson et al., 1984, experiment 1), participants were told that they would have to evaluate different kinds of puzzles designed by psychologists for cognitive testing. In the experimental condition, they were warned that they would have to think of reasons for their evaluations, whereas no such warning was given in the control condition. After they had played with the puzzles for a few minutes, participants had to evaluate the puzzles but, previous to this evaluation, participants in the experimental condition had to spell out their reasons on a blank sheet, ostensibly to 'help them organize their thoughts'. Finally, the experimenter told the participants that they would have to wait for a few minutes but gave them a sheet with new puzzles (of the same kinds as those that had been evaluated) to ease the waiting. By looking at which kinds of puzzles the participants filled in during this free time, it was possible to measure their attitude-behaviour consistency: were they actually doing the puzzles that they said they liked? It turned out that while there was a significant relationship between the evaluations of the different puzzles and the time spend solving them in the control condition, the correlation became non-significant in the experimental condition, when participants had had to state reasons for their evaluations. These results have been replicated many times (see the above references). Clearly, reasoning does not promote a clear understanding of one's preferences. Moreover, the lack of correlation between attitude and behaviour created by too much reasoning can lead participants to form intransitive preferences (Lee, Amir, & Ariely, 2008).

Using similar paradigms in which some participants are asked for reasons, it was found that providing reasons led participants to choose items with which they later were less satisfied (Wilson et al., 1993) or that were less in line with the ratings of experts (McMackin & Slovic, 2000; Wilson & Schooler, 1991). Participants became less good at predicting the results of basketball games (Halberstadt & Levine, 1999). People who think too much are also less apt to understand other people's behaviour (Albrechtsen, Meissner, & Susa, 2009; Ambady, Bernieri, & Richeson, 2000; Ambady & Gray, 2002). This stream of experiments was later followed up by Dijksterhuis and his colleagues who introduced a modified paradigm. Participants are provided with lists of features describing different items (such as flats, cars, etc.) designed in such a way that some items have more positive features. In the baseline condition participants had to provide an answer right 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 performances were obtained in this last condition: unconscious thought was superior to conscious thought (and to immediate decision) (Dijksterhuis, 2004; Dijksterhuis et al., 2006; Dijksterhuis & van Olden, 2006).

Some of Dijksterhuis' results, however, have proven hard to replicate (Acker, 2008; Newell, Wong, Cheung, & Rakow, In Press; Thorsteinson & Withrow, 2009), and alternative interpretations have been offered in some cases (Lassiter, Lindberg, Gonzalez-Vallejo, Bellezza, & Phillips, 2009). In a meta-analysis of this literature, Acker observed that only in a few experiments was unconscious thought significantly superior to conscious thought (Acker, 2008), amounting to a null result when all the experiments were taken into account. Even so, there was no significant advantage of conscious thought over an immediate choice. This is typically the kind of situation in which, according to the classical theories, reasoning should help: a new choice, 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 sub-section we will explore a framework that aims at explaining 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 eighties, a group of leading researchers in decision making elaborated the framework of reason-based choice (Shafir, Simonson, & Tversky, 1993) provides an early review). According to this theory people often make decisions because they can find reasons supporting these decisions. These reasons will not point towards the best decisions or towards decisions that satisfy some criterion of rationality, but towards decisions that can easily be justified and that are less at risk of being criticized. According to the argumentative theory, this is what should happen when people are faced with decisions for which they only have weak intuitions. In this case, reasoning can be used to tip the scales in favour of the choice for which reasons are most easily available. One will thus be at least able to defend the decision if its outcome proves to be unsatisfactory.

Reason based choice is paradigmatically well illustrated in a landmark article of Simonson (Simonson, 1989) in which he studied in particular the attraction effect (Huber, Payne, & Puto, 1982) (see Briley, Morris, & Simonson, 2000) for a cross-cultural variation). The attraction effect occurs when, to a set of two equally valuable alternatives, is added a third alternative that is asymmetrically dominated. This addition tends to increase the rate of choice for the dominating option, in a manner not warranted by rational models (more precisely, this violates the regularity principle (Luce, 1977). Here is one example used in Simonson's experiments. Participants had to choose between packs of beer that varied along the two following dimensions: price and quality. Beer A was of lower quality than beer B but was cheaper and the two attributes balanced in such a way that both beers were regularly chosen in a direct comparison. Some participants, however, 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 but cheaper than this other beer!"). In order to confirm this intuition, Simonson made and tested the three following predictions: (i) a choice based on reasons should be reinforced when participants have to justify themselves; (ii) a choice based on reasons will be perceived as easier to justify and less likely to be criticized; and (iii) a choice based on reasons should give rise to more elaborate explanations. The results of three experiments supported these predictions: participants who had to justify themselves were more likely to choose the dominating option (again in the case of the attraction effect), this choice was seen as being easier to justify, and it was supported by more verbose explanations. Moreover, results also showed that participants who made choices based on reasons tended to make choices that fitted less with their own preferences as stated before the choice. Finally, another set of experiments demonstrated that when participants were more able to use their intuitions, 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, Shocker, & Stewart, 1987). Several well known challenges to the view of humans as making rational decision thanks to their reasoning abilities have been or can be reinterpreted as cases of reason-based choices.

5.3 What reason-based choice can explain

5.3.1 Disjunction effect

The sure thing principle (Savage, 1954) states that if someone favours 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 incertitude regarding E. Shafir and Tversky have recorded several violations of this principle (Shafir & Tversky, 1992; Tversky & Shafir, 1992). 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 [won $200/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.

In both the win and the lose versions 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 supporting 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 now choose to reject the second gamble, even though they would have accepted it for any result of the first gamble. The authors further tested this explanation by devising a comparison that had the same properties as the one above expect for the fact that the reasons supporting the 'accept' decision were the same irrespective of the outcome of the first gamble:

Imagine that you have just played a game of chance that gave you a 50% chance to win $600 and a 50% chance to win $300. Imagine that the coin has already been tossed, but that you will not know whether you have won $600 or $300 until you make your decision concerning a second gamble: 50% chance to win $200 and 50% chance to lose $100.

Here, participants can always justify their decision to accept the second gamble by their previous windfall and, as a result, they make exactly the same choices irrespective of whether the outcome of the first gamble is known or not (see Croson, 1999, for a similar experiment with a variant of the prisoner's dilemma).

5.3.2 Sunk costs 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). It can be illustrated with the following problem (ibid, experiment 3):

As the president of an airline company, you have invested 10 million dollars of the company's money into a research project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed, another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company is building. The question is: should you invest the last 10% of the research funds to finish your radar-blank plane? Yes No

Faced with such a problem, 84% of the participants choose to invest the remaining 10%, even though the project will clearly enough never be profitable. A real life analogue to this problem gave its name to the same phenomenon in the animal behaviour literature where it is known as the "Concorde fallacy" (Dawkins & Brockmann, 1980). Arkes and Ayton have argued that such mistakes result from an unsatisfactory use of explicit reasons such as 'do not waste' (Arkes & Ayton, 1999). We will quickly review the evidence they presented, and add some more.

First of all, Arkes and Ayton contrast the robust sunk cost effects observed in humans (Arkes & Blumer, 1985; Garland, 1990; Staw, 1981) to the absence of such mistakes among animals[xi]. They also point to the fact that children do not seem to be prone to this error. If reasoning was the cure, and not the cause of this phenomenon, the opposite would be expected. More recent experiments make the same point: the number of children committing the fallacy increases with age, from 5 to 14 year-old (Klaczynski & Cottrell, 2004; Morsanyi & Handley, 2008). Moreover, these authors also observed a positive correlation between the tendency to commit the fallacy and cognitive ability—again, not what one would expect if reasoning was the solution. 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). On the other hand, when participants find it harder to justify changing their course of actions, they are more likely to commit the fallacy (J. D. Bragger, Hantula, Bragger, Kirnan, & Kutcher, 2003; J. L. Bragger, Bragger, Hantula, & Kirnan, 1998).

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). An explanation that can be seen as complementary or alternative is that different frames make reasons more or less available, thus modifying the way reasoning affects our decisions. Several results support this interpretation (see also McKenzie, 2004; McKenzie & Nelson, 2003). First, participants who are high in Need for Uniqueness (Snyder & Fromkin, 1977) will tend, when they are accountable for their answers, to be less influenced by frames (Simonson &

Nowlis, 2000). This is because these participants tend to give unconventional reasons for their decisions. Such a variation would not be expected if answers were based only on intuitions. Second, as mentioned earlier, participants who reason more about the tasks are more influenced by framing effects (Igou & Bless, 2007). Finally, when groups make decisions on framed problems, the group converges on the answer that is supported by the strongest reasons (T. W. McGuire, Kiesler, & Siegel, 1987; Milch, Weber, Appelt, Handgraaf, & Krantz, 2009; Paese, Bieser, & Tubbs, 1993). If the answers of the participants were truly based on their intuitions, the answer of 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

Being able to correctly evaluate preferences is necessary for economic models of decision making. However, 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, Loewenstein, & White, 1992; Irwin, Slovic, Lichtenstein, & McClelland, 1993; Kahneman & Ritov, 1994; Tversky, Sattath, & Slovic, 1988). In an early demonstration of this effect, Slovic asked participants to adjust the value of the attribute of one item so that it would be as valuable as another item (Slovic, 1975). Participants could adjust, for instance, the price of a less good orange juice so that it would have the same overall value as a better one. A week after they had done so, participants were asked to choose one of these two items. They overwhelmingly went for the item that was superior on the dimension thought to be the most important, because “it provides 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).

Measures of preferences can also be deeply affected by whether they are gathered jointly or in isolation (Hsee, 1996b, 1998; Hsee, Loewenstein, Blount, & Bazerman, 1999). For instance, people rate a musical dictionary having 10.000 entries that is ‘like new’ higher than one with 20.000 entries and a torn cover when these are presented in isolation. However, when people have to choose between the two, they favour the one that has more entries, despite the torn cover. Although such effects are typically explained by the differences in ease of evaluation of different attributes, they fit perfectly in the current framework. In the example above, people lack reliable intuitions—they cannot tell how many entries a good musical 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 good or a poor price. Dimensions with numerical values will often provide compelling justifications when options are jointly presented. 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, Zhang, Yu, & Xi, 2003, p.16, see also Okada, 2005). Indeed, ‘rationalistic’ attributes make for easy justifications. For instance, in one experiment participants either had to choose between, or to rate the two following options: a roach shaped chocolate of two ounces worth two dollars and a heart shaped chocolate of half an ounce worth 50 cents (Hsee, 1999). A majority (68%) of participants choose the roach shaped chocolate, even though more than half (54%) thought they would enjoy the other more. The participants who choose the bigger, roach shaped chocolate did it because the feeling of disgust, being ‘irrational’, was hard to justify, especially compared to the difference in price and size. However, in light of the results from the psychology of disgust (e.g., Rozin, Millman, & Nemeroff 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, Heath, & Odean, 2003). People will use a ‘more variety is better’, or a ‘don’t pick the same things as others’ rule 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). The use of a ‘don’t pay for delays’ rule will lead to behaviours that go against one’s own interest (Amir & Ariely, 2003). When forecasting their affective states, people will rely on explicit lay theories (Igou, 2004), theories that 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 an 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, In press). 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, Greifeneder, & Todd, 2009), the feature creep effect (D. V. Thompson, Hamilton, & Rust, 2005), the endowment effect (E. J. Johnson, Haubl, & Keinan, 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 to choose an item (Carpenter, Glazer, & Nakamoto, 1994), and sometimes as a reason to reject it (Simonson, Carmon, & O'Curry, 1994; Simonson, Nowlis, & Simonson, 1993), 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, Lerner, & Boettger, 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 one is likely 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 (D. V. Thompson & Norton, 2008) or when they are giving advice (in which case one has to be able to justify oneself, see (Kray & Gonzalez, 1999). On the contrary, when they are choosing for others rather than for themselves they are less prone to these effects, because there is then less of a need for a utilitarian, justifiable decision (Hamilton & Thompson, 2007).

6 Conclusion: reasoning and rationality

The evidence reviewed here shows not only that reasoning falls quite short of reliably delivering rational beliefs and rational decisions, it may even be, in a variety of cases, detrimental to rationality. Reasoning can lead to poor outcomes not because humans are bad at it but because they systematically strive for arguments to justify their beliefs or their actions. This explains the confirmation bias, motivated reasoning, and reason-based choice. These fundamental biases and flaws of human reasoning are well known. What the argumentative theory does is put them in a different perspective. Human reasoning is not a profoundly flawed general mechanism, it is a remarkably efficient specialized device (Mercier, submitted). It is not specialized for a specific domain of knowledge, but for a certain type of social and cognitive interaction. The task of finding persuasive reasons and accurately assessing others' arguments, in which reasoning excels, is no small matter.

Even from a strictly epistemic point of view, 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 favour of the opinions or decisions of the arguer whatever their soundness, and the evaluation of arguments, which aims at differentiating good arguments from bad ones and thereby 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 search for genuine information but rather consider them from the start as counter-arguments to be rebuked. 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 don't have a 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 favour of a variety of hypotheses most of which or even all of which are false, they concur in recognizing sound arguments. These tasks having a demonstrably valid solution, truth indeed wins. If we generalize to problems that do not have a provable solution, we should expect, if not necessarily truth, at least good arguments to win (and we have reviewed in section 2 evidence that such is indeed the case). This may sound trivial, but it is not. It demonstrates that, contrary to standard 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 after winning a debate.

Couldn't the same type of situation that favours sound evaluation favour comparable soundness in the production of arguments? Note, to begin with, that situations of shared interest in truth which cause participants in a group task to evaluate arguments correctly do not suffice to cause them to produce correct arguments. In these group tasks, individual participants come up and propose to the group the same inappropriate answers that they come up with in individual testing. The group success is due first and foremost to the filtering of a variety of solutions achieved through evaluation. When none of the answers initially proposed is correct, then all of them are likely to be rejected, wholly or partly new hypotheses are likely to be proposed, and again filtered, thus explaining how groups may even do better than any of their individual members.

Individuals thinking on their own without the benefit of the wisdom of others can only assess their own hypotheses, but in so doing, they are both judge and party, or rather judge and advocate, which is not an optimal stance to pursue the truth. Couldn't, in principle, an individual 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 to thereby gain in objectivity. Presumably this is what the 10% or so of people who pass the standard Wason selection task do. Doing so is an acquired skill that consists in exerting some imperfect control over a natural disposition that spontaneously pulls in a different direction.

At this point, 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 the moral domain. This is undeniable, but the achievements involved are all collective and result from interactions over many generations. 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, contributing to a kind of division of cognitive labour. Moreover to excel in such groups involves anticipating the manner in which others may evaluate one's arguments, and adjusting these arguments accordingly. Sometimes it may even cause one to recognize that one's difficulty in successfully arguing for one's hypotheses is due to a flaw in the hypotheses themselves and to revise them accordingly. Such recognition, we have suggested, requires a painstakingly acquired ability to exert some limited control over one's own biases. Even among scientists, such an 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 take their highly visible almost freakish contributions as paradigmatic examples of human reasoning. Generally, one should be careful about using the stiking accomplishments of reasoning as proof of its general 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 can initially guess, 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.

References

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.
Albrechtsen, J. S., Meissner, C. A., & Susa, K. J. (2009). Can intuition improve deception detection performance? *Journal of Experimental Social Psychology, 45*(4), 1052–1055.
Allport, F. (1924). *Social Psychology*. Boston: Houghton Mifflin.
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 M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 32, pp. 201–271). New York: Academic Press.
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*, 947–961.
Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology, 18*, 894-897.
Amir, O., & Ariely, D. (2003). Decision by rules: Disassociation between preferences and willingness to act. *Working paper, Massachusetts Institute of Technology, Cambridge, MA*.
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.
Anderson, T., Howe, C., & Tolmie, A. (1996). Interaction and mental models of physics phenomena: Evidence from dialogues between learners. In J. Oakhill & A. Garnham (Eds.), *Mental Models in Cognitive Science: Essays in Honour of Phil Johnson-Laird* (pp. 247-273). Hove: The Psychology Press.
Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (In Press). Large Stakes and Big Mistakes.
Ariely, D., & Levav, J. (2000). Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of Consumer Research, 27*(3), 279-290.
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.
Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*(1), 124-140.
Arkes, H. R., Guilmette, T. J., Faust, D., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of applied psychology, 73*(2), 305-307.
Bailenson, J. N., & Rips, L. J. (1996). Informal reasoning and burden of proof. *Applied Cognitive Psychology, 10*(7), 3-16.
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), 1636-1652.
Baum, L. A., Danovitch, J. H., & Keil, F. C. (2007). Children's sensitivity to circular explanations. *Journal of Experimental Child Psychology, 100*(2), 146-155.
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), 220-240.
Berger, J. A., & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research, 34*.
Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal Reasoning in Rats. *Science, 311*(5763), 1020-1022.
Blanchette, I., & Dunbar, K. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition, 29*(5), 730-735.
Blinder, A. S., & Morgan, J. (2000). Are two heads better than one?: An experimental analysis of group vs. individual decision making. *NBER Working Paper*.
Blum-Kulka, S., Blondheim, M., & Hacohen, 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), 1569-1594.

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.
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), 240-254.
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*, 719–736.
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.
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), 303-345.
Bragger, J. D., Hantula, D. A., Bragger, D., Kirnan, J., & Kutcher, E. (2003). When success breeds failure: history, hysteresis, and delayed exit decisions. *Journal of Applied Psychology, 88*(1), 6-14.
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*, 229-253.
Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573–604.
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), 157-178.
Brock, T. C. (1967). Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology, 3*(3), 269-309.
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), 372-385.
Brownstein, A. L. (2003). Biased predecision processing. *Psychological Bulletin, 129*(4), 545-568.
Butera, F., Legrenzi, P., Mugny, G., & Pérez, J. A. (1992). Influence sociale et raisonnement. *Bulletin de Psychologie, 45*, 144–154.
Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*.
Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. New York: Oxford University Press.
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.
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*, 7-42.
Carpenter, G. S., Glazer, R., & Nakamoto, K. (1994). Meaningful Brand from Meaningless Differentiation: The Dependence on Irrelevant Attributes. *Journal of Marketing Research, 31*(3), 339-350.
Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford Press.
Chaiken, S., & Trope, Y. (1999). *Dual-Process Theories in Social Psychology*. New York: The Guilford Press.
Chaiken, S., & Yates, S. (1985). Affective-cognitive consistency and thought-induced attitude polarization. *Journal of Personality and Social Psychology, 49*(6), 1470-1481.
Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38*, 191-258.
Chernev, A. (2005). Context effects without a context: Attribute balance as a reason for choice. *Journal of Consumer Research, 32*(2), 213-223.
Clark, R. A., & Delia, J. G. (1976). The development of functional persuasive skills in childhood and early adolescence. *Child Development*, 1008-1014.
Claxton, G. (1997). *Hare Brain, Tortoise Mind: How Intelligence Increases When You Think Less*. New York: Harper Collins.
Clément, F. (In Press). To trust or not to trust? Children's social epistemology. *Review of Philosophy and Psychology*.
Conlon, E. J., & Wolf, G. (1980). The moderating effects of strategy, visibility, and involvement on allocation behavior: An extension of Staw's escalation paradigm. *Organizational Behavior and Human Performance, 26*, 172-192.
Corner, A., Hahn, U., & Oakfsord, M. (2006). The slippery slope argument: probability, utility and category reappraisal. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*.
Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187-276.
Cowley, M., & Byrne, R. M. J. (2005). *When falsification is the only path to truth.* Paper presented at the Twenty-Seventh Annual Conference of the Cognitive Science Society, Stresa, Italy.
Croson, R. T. A. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes, 80*(2), 118-133.
Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infants. *Cognition, 72*, 237-267.
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.
Davis, J. H. (1973). Group decisions and social interactions: A theory of social decision schemes. *Psychological Review, 80*, 97-125.
Dawkins, R., & Brockmann, H. J. (1980). Do digger wasps commit the Concorde fallacy. *Animal Behaviour, 28*, 892-896.
Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (pp. 282-309). Oxford: Basil Blackwell Scientific Publications.
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.
Dehaene, S. (1999). *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.
Dennett, D. C. (1969). *Content and Consciousness*. London: Routledge and Kegan Paul.
Dijksterhuis, A. (2004). Think different: the merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87*(5), 586-598.
Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*(5763), 1005-1007.
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), 627-631.
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), 568-584.
Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, 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.

Ditto, P. H., Scepansky, 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.

Dunbar, R. I. M. (1996). The social brain hypothesis. *Evolutionary Anthropology, 6*, 178–190.

Dunbar, R. I. M., & Shultz, S. (2003). Evolution of the social brain. *Science 302*, 1160-1161.

Dunn, J., & Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology, 23*, 791-798.

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), 1082-1090.

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.

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), 316-327.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71*, 5-24.

Esser, J. K. (1998). Alive and well after 25 years: A review of groupthink research. *Organizational Behavior and Human Decision Processes, 73*(2–3), 116–141.

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).

Evans, J. S. B. T. (1989). *Bias in Human Reasoning: Causes and Consequences*. Hillsdale, NJ: Lawrence Erlbaum.

Evans, J. S. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology, 87*, 223-240.

Evans, J. S. B. T. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychological bulletin, 128*(6), 978-996.

Evans, J. S. B. T. (2007). *Hypothetical Thinking: Dual Processes in Reasoning and Judgment*. Hove: Psychology Press.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition, 11*, 295-306.

Evans, J. S. B. T., & Frankish, K. (Eds.). (2009). *In Two Minds*. Oxford: Oxford University Press.

Evans, J. S. 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), 1495-1513.

Evans, J. S. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*(3), 391-397.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human Reasoning: The Psychology of Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and Reasoning*. Hove: Psychology Press.

Evans, J. S. B. T., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology, 67*, 479-486.

Farnsworth, P. R., & Behner, A. (1931). A note on the attitude of social conformity. *Journal of Social Psychology, 2*, 126-128.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives, 10*(110–118).

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition, 77*(1), 1-79.

Foot, H., Howe, C., Anderson, A., Tolmie, A., & Warden, D. (1994). *Group and Interactive Learning*. Southampton: Computational Mechanics Press.
Fox, F. V., & Staw, B. M. (1979). The trapped administrator: The effects of job insecurity and policy resistance upon commitment to a course of action. *Administrative Science Quarterly, 24*, 449-471.
Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., et al. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology, 92*(2), 325.
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), 728-731.
Geurts, B. (2003). Reasoning with quantifiers. *Cognition, 86*(3), 223-251.
Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.
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), 503-514.
Gilovich, T. (1983). Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology, 44*(6), 1110-1126.
Girotto, V., Kemmelmeier, 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.
Glachan, M., & Light, P. (1982). Peer interaction and learning: Can two wrongs make a right? In G. Butterworth & P. Light (Eds.), *Social cognition: Studies in the development of understanding* (pp. 238–262). Chicago: University of Chicago Press.
Gladwell, M. (2005). *Blink: The Power of Thinking without Thinking*. Boston: Little, Brown, .
Green, K. C., Armstrong, J. C., & Graefe, A. (2007). Methods to elicit forecasts from groups delphi and prediction markets compared. *MPRA Paper No. 4663*.
Greenwald, A. G. (1969). The open-mindedness of the counterattitudinal role player. *Journal of Experimental Social Psychology, 5*(4), 375-388.
Griffin, D. W., & Dunning, D. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality, 59*(6), 1128-1139.
Guenther, C. L., & Alicke, M. D. (2008). Self-enhancement and belief perseverance. *Journal of Experimental Social Psychology, 44*(3), 706-712.
Hagler, D. A., & Brem, S. K. (2008). Reaching agreement: The structure & pragmatics of critical care nurses' informal argument. *Contemporary Educational Psychology, 33*(3), 403-424.
Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review, 114*(3), 704-732.
Hahn, U., Oaksford, M., & Bayindir, H. (2005). How convinced should we be by negative evidence? *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*.
Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.
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), 517-530.
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), 546-555.
Harris, P. L. (2007). Trust. *Developmental Science, 10*, 135-138.
Hart, W., Albarracin, D., Eagly, A. H., Lindberg, M., Merrill, L., Brechan, I., et al. (in press). Feeling validated versus being correct? A meta-analysis of selective exposure to information. *Psychological Bulletin*.
Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91*, 517-539.

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), 116-126.
Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69*, 1069-1086.
Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 719-731.
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), 459-475.
Hrdy, S. B. (2009). *Mothers and Others*. Cambridge, MA: Belknap Press.
Hsee, C. K. (1995). Elastic justification: How tempting but task-irrelevant factors influence decisions. *Organizational Behavior and Human Decision Processes, 62*(3), 330-337.
Hsee, C. K. (1996a). Elastic Justification: How Unjustifiable Factors Influence Judgments. *Organizational Behavior and Human Decision Processes, 66*(1), 122-129.
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), 247-257.
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*.
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 and Review, 6*(4), 555-561.
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.
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), 576-590.
Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology, 86*(5).
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), 257-272.
Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *The Journal of Consumer Research, 9*(1), 90-98.
Humphrey, N. K. (1976). The social function of Intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing Points in Ethology* (pp. 303-317). Cambrige, Massachusetts: Cambridge University Press.
Igou, E. R. (2004). Lay theories in affective forecasting: The progression of affect. *Journal of Experimental Social Psychology, 40*(4), 528-534.
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.
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.
Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology, 50*(6), 1141-1151.
Janis, I. L. (1982). *Groupthink* (2nd Rev. ed.). Boston: Houghton Mifflin.
Janis, I. L., & Mann, L. (1977). *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment*. New York: Free Press.
Jellison, J. M., & Mills, J. (1969). Effect of public commitment upon opinions. *Journal of Experimental Social Psychology, 5*(3), 340-346.
Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, UK: Cambridge University Press.
Johnson-Laird, P. N. (2006). *How We Reason*. Oxford: Oxford University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646-678.

Johnson-Laird, P. N., & Wason, P. C. (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology, 22*(1), 49-61.

Johnson, B. T., & Eagly, A. H. (1989). Effects of involvement on persuasion: A meta-analysis. *Psychological Bulletin, 106*(2), 290-314.

Johnson, D. W., & Johnson, R. (2007). *Creative constructive controversy: Intellectual challenge in the classroom* (4th ed.). Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher, 38*(1), 37.

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*(3), 461-473.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision, 50*(1), 59-99.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49-81). Cambridge, UK: Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267–294). Cambridge, UK: Cambridge Univiversity Press.

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.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430-454.

Kaplan, M. F., & Miller, C. E. (1977). Judgments and group discussion: Effect of presentation and memory factors on polarization. *Sociometry, 40*(4), 337-343.

Katz, J. J. (1986). *Cogitations*. New York: Oxford University Press.

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.

Kerr, N. L., Maccoun, R. J., & Kramer, G. P. (1996). Bias in judgement: comparing individuals and groups. *Psychological review, 103*(4), 687-719.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*, 623-655.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271-304.

Klaczynski, P. A. (1997). Bias in adolescents' everyday reasoning and its relationship with intellectual ability, personal theories, and self-serving motivation. *Developmental Psychology, 33*, 273-283.

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), 147-174.

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), 2873-2891.

Klaczynski, P. A., & Gordon, D. H. (1996b). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology, 62*, 317-339.

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*, 470-485.

Klaczynski, P. A., & Lavallee, 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*, 1-24.

Klaczynski, P. A., & Narasimham, G. (1998). Development of scientific reasoning biases: Cognitive versus ego-protective explanations. *Developmental Psychology, 34*, 175-187.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning tasks. *Psychology and Aging, 15*, 400-416.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychol Rev, 107*(4), 852-884.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Klein, G. (1998). *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.

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.

Kogan, N., & Wallach, M. A. (1966). Modification of a judgmental style through group interaction. *Journal of Personality and Social Psychology, 4*(2), 165-174.

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), 669-685.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory and Cognition, 6*, 107-118.

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).

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2ème ed., pp. 390-402). Oxford: Basil Blackwell Scientific Publications.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19*(5), 448-468.

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

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(22), 155-178.

Kuhn, D., & Lao, J. (1996). Effects of Evidence on Attitudes: Is Polarization the Norm? *Psychological Science, 7*, 115-120.

Kuhn, D., Shaw, V. F., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15*, 287-315.

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*, 289–296.

Kunda, Z. (1987). Motivation and inference: Self-serving generation and evaluation of evidence. *Journalof Personality and Social Psychology, 53*(636-647).

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498.

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), 437-459.

Lao, J., & Kuhn, D. (2002). Cognitive engagement and attitude development. *Cognitive Development, 17*(2), 1203-1217.

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, 9999*(9999).

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88*, 605-620.
Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22*, 177–189.
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*, 644–651.
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*, 50-67.
Laughlin, P. R., Zander, M. L., Knievel, 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*, 684-694.
Lee, L., Amir, O., & Ariely, D. (2008). In search of Homo economicus: Preference consistency, emotions, and cognition. . *Unpublished Working Paper. Columbia University.*
Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*, 255-275.
Leslie, A. M. (1987). Pretense and representation: The origins of a "theory of mind" • . *Psychological Review, 94*(4), 412-426.
Leslie, A. M. (1994). ToMM, ToBY and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind*. Cambridge: Cambridge University Press.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*(3), 265-288.
Liberman, A., & Chaiken, S. (1991). Value conflict and thought-induced attitude change. *Journal of Experimental Social Psychology, 27*(3), 203-216.
Littlepage, G. E., & Mueller, A. L. (1997). Recognition and utilization of expertise in problem-solving groups: Expert characteristics and behavior. *Group Dynamics, 1*, 324-328.
Lombardelli, C., Proudman, J., & Talbot, J. (2005). Committees versus individuals: An experimental analysis of monetary policy decision-making. *International Journal of Central Banking, May*, 181-205.
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), 2098-2109.
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*, 35-66.
Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 15*(3), 215-233.
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), 854-870.
Madsen, D. B. (1978). Issue importance and group choice shifts: A persuasive arguments approach. *Journal of Personality and Social Psychology, 36*, 1118-1127.
Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*(2), 161-175.
Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*, 367–380.
Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science, 19*(3), 255-260.
Maynard-Smith, J. (1994). Must reliable signals always be costly? *Animal Behaviour, 47*, 1115-1120.

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), 917-930.
McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1). New York: Academic Press.
McKenzie, C. R. M. (2004). Framing effects in inference tasks—and why they're normatively defensible. *Memory & cognition, 32*, 874–885.
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 and Review, 10*(3), 596-602.
McMackin, J., & Slovic, P. (2000). When does explicit justification impair decision making? *Journal of Applied Cognitive Psychology, 14*, 527–541.
Mercier, H. (In prep). Developmental evidence for the argumentative theory of reasoning.
Mercier, H. (submitted). Looking for arguments.
Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferences. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds*. New York: Oxford University Press.
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), 834-839.
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*.
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), 259-269.
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*, 189-202.
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), 561-574.
Molden, D. C., & Higgins, E. T. (2005). Motivated thinking. In K. Holyoak & R. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning*. Cambridge Cambridge University Press.
Moorhead, G., Ference, R., & Neck, C. P. (1991). Group decision fiascoes continue: Space shuttle Challenger and a revised groupthink framework. *Human Relations, 44*(6), 539.
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.
Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning, 4*(3), 231-248.
Navarro, A. D., & Fantino, E. (2005). The Sunk Cost Effect In Pigeons And Humans. *Journal of the Experimental Analysis of Behavior, 83*(1), 1.
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), 367-380.
Neuman, Y., Weinstock, M. P., & Glasner, A. (2006). The effect of contextual factors on the judgement of informal reasoning fallacies. *The Quarterly Journal of Experimental Psychology, 59*(2), 411-425.
Newell, B. R., Wong, K. Y., Cheung, J. C. H., & Rakow, T. (In Press). Think, blink or sleep on it? The impact of modes of thought on complex decision making. *The Quarterly Journal of Experimental Psychology*.
Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: testing the predictions of theories of syllogistic reasoning. *Memory and Cognition, 27*(2), 344-354.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology, 2*, 175-220.

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences, 12*(7), 265-272.
Nussbaum, E. M. (2008). Collaborative Discourse, Argumentation, and Learning: Preface and Literature Review. *Contemporary Educational Psychology, 33*(3), 15.
Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*(3), 384-395.
Nyhan, B., & Reifler, J. (In prep.). When Corrections Fail.
Oaksford, M., Chater, N., & Grainger, R. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5*, 193-243.
Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology, 58*(2), 75-85.
Okada, E. M. (2005). Justification effects on consumer choice of hedonic and utilitarian goods. *Journal of Marketing Research, 42*(1), 43-53.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old infants understand false beliefs? *Science, 308*, 255-258.
Ormerod, P. (2005). *Why Most Things Fail: Evolution, Extinction and Economics*. London: Faber & Faber.
Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin and Review, 11*(6), 988-1010.
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*, 149-149.
Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision-making. *Cognition, 49*, 123-163.
Perelman, C., & Olbrechts-Tyteca, L. (1969). *The New Rhetoric: A Treatise on Argumentation*. Notre Dame, IN: University of Notre Dame Press.
Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*, 562-571.
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*, 349-360.
Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123-205). Orlando, FL: Academic Press.
Petty, R. E., & Cacioppo, J. T. (1990). Involvement and persuasion: Tradition versus integration. *Psychological Bulletin, 107*(3), 367-374.
Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology* (Vol. 1, pp. 323–390). Boston: McGraw-Hill.
Piaget, J. (1928). *Judgment and Reasoning in the Child*. London: Routledge and Kegan Paul.
Poletiek, F. H. (1996). Paradoxes of falsification. *Quarterly Journal of Experimental Psychology, 49A*, 447-462.
Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology, 69*(3), 408-419.
Powell, C. (2003). The Delphi technique: myths and realities. *Journal of Advanced Nursing, 41*(4), 376-382.
Prasad, M., Perrin, A. J., Bezila, K., Hoffman, S. G., Kindleberger, K., Manturuk, K., et al. (2009). " There must be a reason": Osama, Saddam, and inferred justification.
Pritchard, D. (2005). *Epistemic Luck*. New York: Clarendon Press.
Pyszczynski, T., & Greenberg, J. (1987). Toward and integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.),

Advances in Experimental Social Psychology (Vol. 20, pp. 297-340). New York: Academic Press.
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.
Redlawsk, D. P. (2002). Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making. *The Journal of Politics, 64*(4), 1021-1044.
Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction, 11*(3/4), 347-364.
Ricco, R. B. (2003). The macrostructure of informal arguments: A proposed model and analysis. *The Quarterly Journal of Experimental Psychology A, 56*(6), 1021-1051.
Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, MA: MIT Press.
Rips, L. J. (1998). Reasoning and conversation. *Psychological Review, 105*, 411–441.
Rips, L. J. (2002). Circular reasoning. *Cognitive Science, 26*, 767–795.
Roberts, M. J., & Newton, E. J. (2002). Inspection times, the change task, and the rapid response selection task. *Quarterly Journal of Experimental Psychology, 54*, 1031-1048.
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), 880-802.
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), 627-634.
Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting, 15*(4), 353-375.
Rozin, P., Millman, L., & Nemeroff , C. (1986). Operation of the laws of sympathetic magic in disgust and other domain. *Journal of Personality and Social Psychology, 50*(4), 703-712.
Russo, J. E., Carlson, K. A., & Meloy, M. G. (2006). Choosing an inferior alternative. *Psychological Science, 17*(10), 899-904.
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), 1149-1161.
Sacco, K., & Bucciarelli, M. (2008). The role of cognitive and socio-cognitive conflict in learning to reason. *Mind & Society, 7*(1), 1-19.
Sadler, O., & Tesser, A. (1973). Some effects of salience and time upon interpersonal hostility and attraction during social isolation. *Sociometry, 36*(1), 99-112.
Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology, 59*(2), 229-241.
Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2009). What moderates the too-much-choice effect? *Psychology and Marketing, 26*(3).
Scholnick, E. K., & Wing, C. S. (1991). Speaking deductively: Preschoolers' use of If in conversation and in conditional inference'. *Developmental Psychology, 27*(2), 249–258.
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), 1080-1093.
Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs make a right. . .if they argue together! *Cognition and Instruction, 18*, 461–494.
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.
Sengupta, J., & Fitzsimons, G. J. (2000). The effects of analyzing reasons for brand preferences: Disruption or reinforcement? *Journal of Marketing Research, 37*(3), 318-330.

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), 705-711.

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

Shafir, E., & Tversky, A. (1992). Thinking through Uncertainty: Nonconsequential Reasoning and Choice. *Cognitive Psychology, 24*(4), 449-474.

Shaw, V. F. (1996). The cognitive processes in informal reasoning. *Thinking & Reasoning, 2*(1), 51-80.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

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

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research, 27*(2), 150-162.

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*, 23-23.

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.

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.

Slavin, R. E. (1995). *Cooperative Learning: Theory, Research, and Practice* (2nd ed.). London: Allyn and Bacon.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology (Human Perception and Performance), 1*(3), 280-287.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., et al. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323*(5910), 122.

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

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational behavior and human decision processes(Print), 43*(1), 1-28.

Snyder, C. R., & Fromkin, H. L. (1977). Abnormality as a Positive Characteristic: The Development and Validation of a Scale Measuring Need for Uniqueness. *Journal of Abnormal Psychology, 86*(5), 518-527.

Soman, D., & Cheema, A. (2001). The Effect of Windfall Gains on the Sunk-Cost Effect. *Marketing Letters, 12*(1), 51-62.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89-96.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind and Language, 12*(1), 67-83.

Sperber, D. (2000a). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 117-137). Oxford: Oxford University Press.

Sperber, D. (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics, 29*, 401-413.

Sperber, D. (Ed.). (2000b). *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31-95.

Sperber, D., & Mercier, H. (In press). Reasoning as a social competence. In J. Elster & H. Landemore (Eds.), *Collective Wisdom*.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language, 17*, 3-23.

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

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*, 25-35.

Staw, B. M. (1981). The escalation of commitment to a course of action. *Academy of Management Review, 6*(4), 577-587.

Stein, N. L., & Albro, E. R. (2001). The origins and nature of arguments: Studies in conflict understanding, emotion, and negotiation. *Discourse Processes, 32*(2&3), 113-133.

Stein, N. L., & Bernas, R. (1999). The early emergence of argumentative knowledge and skill. In J. Andriessen & P. Coirier (Eds.), *Foundations of Argumentative Text Processing* (pp. 97-116). Amsterdam: Amsterdam University Press.

Stein, N. L., Bernas, R. S., & Calicchia, D. J. (1997). Conflict talk: Understanding and resolving arguments. In T. Givon (Ed.), *Conversation: Cognitive, Communicative and Social Perspectives*. Amsterdam: John Benjamins.

Stein, N. L., Bernas, R. S., Calicchia, D. J., & Wright, A. (1995). Understanding and resolving arguments: The dynamics of negotiation. In B. Britton & A. G. Graesser (Eds.), *Models of Understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Stein, N. L., & Miller, C. A. (1993). The development of meaning and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 4, pp. 285-335). Hillsdale: Lawrence Erlbaum Associates.

Steiner, I. D. (1972). *Group processes and productivity*: New York: Academic Press.

Strahan, E. J., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology, 38*(6), 556-568.

Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy, 10*(2), 175-195.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of Beliefs by 13-Month-Old Infants. *Psychological Science, 18*(7), 580-586.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755-769.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Tesser, A. (1976). Attitude polarization as a function of thought and reality constraints. *Journal of Research in Personality, 10*(2), 183-194.

Tesser, A., & Conlee, M. C. (1975). Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology, 31*(2), 262-270.

Tesser, A., & Leone, C. (1977). Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology, 13*(4), 340-356.

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*, 639-652.

Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology, 57*(3), 388-398.

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), 915-934.

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), 632-640.

Thagard, P. (2005). Testimony, credibility, and explanatory coherence. *Erkenntnis, 63*, 295-316.

Thompson, D. V., Hamilton, R. W., & Rust, R. T. (2005). Feature fatigue: When product capabilities become too much of a good thing. *Journal of Marketing Research, 42*(4), 431-442.
Thompson, D. V., & Norton, M. I. (2008). The social utility of feature creep. In A. Lee & D. Soman (Eds.), *Advances in Consumer Research* (Vol. 35, pp. 181-184). Duluth, MN: Association for Consumer Research.
Thompson, V. A., Evans, J. S. B. T., & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory and Language, 53*(2), 238-257.
Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. (2005). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin and Review, 10*(1), 184-189.
Thorsteinson, T. J., & Withrow, S. (2009). Does unconscious thought outperform conscious thought on complex decisions? A further examination. *Judgment and Decision Making, 4*(3), 235-247.
Tichy, G. (2004). The over-optimism among experts in assessment and foresight. *Technological Forecasting & Social Change, 71*(4), 341-363.
Tindale, R. S., & Sheffey, S. (2002). Shared information, cognitive load, and group memory. *Group Processes & Intergroup Relations, 5*(1), 5.
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.
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), 675-691.
Trognon, A. (1993). How does the process of interaction work when two interlocutors try to resolve a logical problem? *Cognition and Instruction, 11*(3&4), 325-345.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.
Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293-315.
Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*(3), 371-384.
Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*(5), 305-309.
Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., et al. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32*(1), 109-123.
van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311-330.
Vinokur, A. (1971). Review and theoretical analysis of the effects of group processes upon individual and group decisions involving risk. *Psychological Bulletin, 76*(4), 231-250.
Vinokur, A., & Burnstein, E. (1978). Depolarization of attitudes in groups. *Journal of Personality and Social Psychology, 36*(8), 872-885.
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129-137.
Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology: I* (pp. 106–137). Harmandsworth, England: Penguin.
Wason, P. C., & Evans, J. S. B. T. (1975). Dual processes in reasoning? *Cognition, 3*, 141-154.
Webb, N. M., & Palinscar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 841–873). New York: Prentice Hall.
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*, 516-523.

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.
Whiten, A., & Byrne, R. W. (Eds.). (1997). *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge: Cambridge University Press.
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.
Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123-205). Orlando, FL: Academic Press.
Wilson, T. D., Kraft, D., & Dunn, D. S. (1989). The disruptive effects of explaining attitudes: the moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology, 25*(5), 379-400.
Wilson, T. D., & LaFleur, S. J. (1995). Knowing what you'll do: effects of analyzing reasons on self-prediction. *J Pers Soc Psychol, 68*(1), 21-35.
Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101-126.
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.
Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Thinking, 60*(2), 181-192.
Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks, 11*(7-8), 1317-1329.
Wynn, K. (1992). Addition and subtraction in human infants. *Nature, 358*, 749-750.
Xu, J., & Schwarz, N. (In press). Do we really need a reason to indulge. *Journal of Marketing Research*.
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.
Zahavi, A., & Zahavi, A. (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford: Oxford University Press.

---

[i] Recently, 'reasoning' has been used as a mere synonym of inference and is then unproblematically attributed to infants (Spelke & Kinzler, 2007) or to non-human animals (Blaisdell, Sawa, Leising, & Waldmann, 2006). In this article, however we use 'reasoning' in its more common and narrower sense.

[ii] For an elaborate view of coherence and its possible role in cognition and communication, see Thagard (2005).

[iii] At least as far as their initial answer is concerned. Given that participants feel pressured to give a correct answer, they will later use reasoning to try to check their answers. Reasoning being fundamentally biased, this will only lead them to find confirming arguments for their initial answer in most cases (see section 3)

[iv] In the psychology of reasoning, some tasks can be called 'production tasks' because participants have to produce a logically valid conclusion from a set of premises. These tasks are very different from the production of arguments in a debate though. 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.

[v] Other results, slightly weaker, are obtained for inductive tasks (Laughlin, Bonner, & Miner, 2002; Laughlin, Hatch, Silver, & Boh, 2006; Laughlin, VanderStoep, & Hollingshead, 1991; Laughlin, Zander, Knievel, & Tan, 2003). Debates are also a well known manner to improve comprehension in many domains; see for instance (Anderson, Howe, Soden, Halliday, & Low, 2001; Anderson, Howe, & Tolmie, 1996; Foot, Howe, Anderson, Tolmie, & Warden, 1994; Howe, 1990; D. W. Johnson & Johnson, 2007; D. W. Johnson & Johnson, 2009; Lao & Kuhn, 2002; Nussbaum, 2008; Nussbaum & Sinatra, 2003; Slavin, 1995; M. K. Smith et al., 2009; Tolmie, Howe, Mackenzie, & Greer, 1993; van Boxtel, van der Linden, & Kanselaar, 2000; Webb & Palinscar, 1996).

[vi] For the groups whose members had to solve the task individually first. When groups where directly confronted with the task, they 'only' reached 70% of correct answers.

[vii] It should be noted that given the limited number of participants (N=16 per condition), this difference was not statistically significant, despite the fact that participants were more than four times more likely to falsify someone else's hypothesis than their own.

[viii] The Delphi technique is a method of forecasting that can be seen as trying to make the best of the confirmation bias. Its effectiveness shows that in an appropriate context the confirmation bias can be conducive to very good performances (Green, Armstrong, & Graefe, 2007; Keeney, Hasson, & McKenna, 2001; Powell, 2003; Rowe & Wright, 1999; Tichy, 2004).

[ix] It should be noted that 'motivated', or 'motivation' as used here does not refer to a 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.

[x] Attitude polarization is most likely to occur for individuals holding very strong attitude with high confidence. The problem is then that these individuals will tend to fall on one end of the attitude scale before reading the arguments, making it close to impossible to detect any move towards a more extreme attitude.

[xi] Pigeons have recently been shown to fall prey to the fallacy, but only when it was not indicated 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 Arkes and Ayton's point.