Reasoning as a Social Competence

Dan Sperber

Institut Jean Nicod (CNRS/EHESS/ENS)

Hugo Mercier

Philosophy, Politics and Economics Program (University of Pennsylvania)

To be published in:

Collective Wisdom

Landemore, H. and Elster, J. (Eds.)

Groups do better at reasoning tasks than individuals, and, in some cases, do even better than *any* of their individual members. Here is an illustration. In the standard version of Wason selection task (Wason, 1966), the most commonly studied problem in the psychology of reasoning, only about 10% of participants give the correct solution, even though it can be arrived at by elementary deductive reasoning.[1] Such poor performance begs for an explanation, and a great many have been offered. What makes the selection task relevant here is that the difference between individual and group performance is striking. Moshman & Geil for instance (1998) had participants try and resolve the task either individually or in groups of five or six participants. While, unsurprisingly, only 9% of the participants working on their own found the correct solution, an astonishing 70% of the groups did. Moreover, when groups were formed with participants who had first tried to solve the task individually, 80% of the groups succeeded, including 30% of the groups in which none of the members had succeeded on his or her own. How are such differences between individual and group performance to be explained?

Reasoning is quite generally viewed as an ability aimed at serving the reasoner's own cognitive goals. If so, the contribution of reasoning to 'collective wisdom'—for instance, to the collective discovery of the correct solution to the Wason selection task in the study of Moshman & Geil—should be seen as a side-effect, a by-product of its proper function. We want to argue, on the contrary, that the function of reasoning is primarily social and that it is the individual benefits that are side-effects. The function of reasoning is to produce arguments in order to convince others and to evaluate arguments others use in order to convince us. We will show how this view of reasoning as a form of social competence correctly predicts both good and bad performance in the individual and in the collective case, and helps explain a variety of psychological and sociological phenomena.

It is easy enough to adduce examples of impressive, mediocre, or abysmal collective performance: think of teamwork in science, television games such as 'Family Feud', and

[1] In the standard Wason selection task (of which there are a great many variants), participants are presented with four cards that have a number on one side and a letter on the other. Two of the cards are shown with the number side up displaying, say, a 4 and a 7, and two with the letter side up displaying, say, an A and a K. Participants are asked: "Which of these four cards is it necessary to turn over in order to ascertain whether the following claim is true or false: If a card has a vowel on one side, it has an even number on the other side?" Whereas the correct answer is 'A and 7', most of the participants answer 'A' or 'A and 4'. (for an analysis and explanation of the task consistent with the approach to reasoning presented here, see Sperber, Cara, & Girotto, 1995).

lynching. One might explain this variety of cases in different ways. The simplest explanation —but clearly an inadequate one—would be that, in each case, performance results from the aggregation of the contributions of individuals endowed to various degrees with an all-purpose general intelligence or, if you prefer, 'rationality', and that the differences in the quality of the outcomes is what you should expect assuming a normal distribution (the best and worst examples being picked at both end of the bell-shaped curve). A somewhat more plausible explanation would take into account the institutional articulation of individual contributions and help explain the fact that collective performances of a given type (e.g. scientific work vs. mob 'justice') tend to cluster at one or another end of the distribution. In a typical social science fashion, such explanations involve idealized assumption about rationality but no serious consideration of actual mental mechanisms.

It would be preferable, for the sake of simplicity, if a sophisticated understanding of social phenomena could be achieved with little or no psychology, but, we would argue (see Sperber, 2006), this is as implausible as achieving a deep understanding of epidemiological phenomena without a serious interest in pathology—and for similar reasons. We explore rather the possibility that explaining different kinds of collective cognitive performances requires paying attention to the individual psychological mechanisms and dispositions involved.

The common view of human thinking as a relatively homogenous process governed by intelligence or reason and interfered with by passions is based on conscious access to our thoughts and on our ability to integrate our conscious thoughts in discursive form. Conscious access to our thoughts, however, tells us little about thinking proper, that is, about the processes through which we produce these thoughts. The discursive integration of our thoughts tells us little about their articulation in the mind. Empirical research in cognitive psychology strongly suggests that, actually, conscious access to thought *processes* is extremely poor; that there is no unified, domain-general mechanism to which 'reason' or 'intelligence' would refer; that thought processes are carried out by a variety of autonomous mental mechanisms (often described as 'modules'); and that many of these mechanisms use as input or produce as output a variety of intermediate-level mental representations that are not accessible to consciousness (see e.g. Dennett, 1991; Marr, 1982; Sperber, 2001).

Is reasoning, then, the output of a single mental mechanism or of several, and if so of which? In psychology of reasoning, the search had initially been for *the* underlying mechanism, in the singular. The most debated question was, did this mechanism use logical rules (Rips, 1994), mental models (Johnson-Laird, 1983), or pragmatic schemas (Cheng & Holyoak, 1985)?

More recently, many researchers have argued that reasoning can be carried out through two distinct cognitive systems. System 1 processes are typically described as unconscious, implicit, automatic, associative or heuristic. They are seen as fast, cheap, and generally efficient in ordinary circumstances, but prone to mistakes when the conditions or the problems are non-standard. System 2 processes are described on the contrary as conscious, explicit, rule-based, or analytic. They are seen as slow and effortful, but as more systematically reliable and as better at handling non-trivial cases. Actually, such ‘dual system theories’ according to which mental processes can be divided into two broad types are common or even dominant in many fields of psychology. Within cognitive psychology, they started in the fields of attention (Posner & Snyder, 1975) and memory (Schacter, 1987), soon followed by learning (Berry & Dienes, 1993; Reber, 1993), before expanding towards reasoning (Evans & Over, 1996; Sloman, 1996; Stanovich, 2004) and decision making (Kahneman, 2003; Kahneman & Frederick, 2002). They are present in nearly every domain of social psychology: most notably in persuasion and attitude change (Chaiken, Liberman, & Eagly, 1989; Petty & Cacioppo, 1986), but also in attitudes (Wilson, Lindsey, & Schooler, 2000), stereotypes (Bodenhausen, Macrae, & Sherman, 1999) or person perception (Uleman, 1999). They are also found in moral (Haidt, 2001) and developmental (Klaczynski & Lavallee, 2005) psychology.

A good illustration of the speed and apparent effectiveness of system 1 processes is provided by a study of Todorov et al. (2005) on judgments of competence. Participants were shown for just one second the pictures of two faces unknown to them and were asked, which of the two individuals looked more competent. One might think that, in order to judge an individual’s competence, a variety of evidence would be needed and that facial appearance would be of limited relevance. Still, participants showed no qualms answering the question. This reliance on a quasi immediate first impression is a perfect example of system 1 inference. Actually, the faces were those of candidates who had competed for election to the American Senate. Participants' answer predicted the result of the elections with 67.6% accuracy. As the authors note:

> "Actual voting decisions are certainly based on multiple sources of information other than inferences from facial appearance. Voters can use this additional information to modify initial impressions of political candidates. However, from a dual-system perspective, correction of intuitive system 1 judgments is a prerogative of system 2 processes that are attention-dependent and are often anchored on intuitive system 1 judgments. Thus, correction of initial impressions may be insufficient. In the case of voting decisions, these decisions can be anchored on initial inferences of competence from facial appearance. From this perspective, in the absence of any other information, voting preferences should be closely related to such inferences. In real-life voting decisions, additional information may weaken the relation between inferences from faces and decisions but may not change the nature of the relation" (Todorov et al., 2005, 1625)

The Wason selection task (see above) provides the most commonly used example of fast and unconscious system 1 mental processes yielding an answer that happens to be false. Another clear and simpler example is provided by the bat and ball problem (Frederick, 2005): "A bat and a ball together cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?" For most of us, when we are first presented with the problem, an answer springs to mind: The ball costs 10 cents. Presumably, this answer comes somehow from an instantaneous realization that the first amount mentioned ($1.10) equals the second amount mentioned ($1.00) plus 10 cents and from the implicit and unexamined presupposition that this directly provides an answer to the problem. We are able however, but with more effort, to understand that, if the ball cost 10 cents and therefore the bat one dollar, the difference between the two items would not be one dollar as stipulated, but 90 cents. Our initial answer cannot be right! More careful 'system 2' reasoning reveals the correct answer: the ball costs 5 cents (and the bat $1.05, with a total of $1.10 and a difference of one dollar). These two successive answers illustrate the duality of processes involved.

While much evidence has accumulated in favour of a dual system view of reasoning (Evans, 2003, 2008), the contrast between the two postulated systems is left vague, and explanations of why the mind should have such a dual organisation are at best very sketchy. We have suggested a more explicit and principled distinction between ‘intuitive’ and ‘reflective’ inferences (Mercier & Sperber, 2009) that can be seen as a particular version of dual systems theories provided they are broadly characterized, or else as an alternative to these theories, drawing on much of the same evidence and sharing several fundamental hunches. We argue that ‘system 1’ or intuitive inferences are carried out by a variety of domain-specific mechanisms. Reflective inference, which corresponds to reasoning in the ordinary sense of the term, is, we claim, the indirect output of a single module. A distinctive feature of our approach, relevant to the discussion of ‘collective wisdom’, is the claim that the main function of reflective inference is to produce and evaluate arguments occurring in interpersonal communication (rather than to help individual ratiocination).

Intuitive inference

In dual system theories of reasoning, ‘reasoning’ and ‘inference’ are used more or less as synonyms. We prefer to use ‘inference’ in a wider sense common in psychology, and ‘reasoning’ in a narrower sense common in ordinary language and in philosophy. An inference, as the term is used in psychology, is a process that, given some input information reliably yields as output further information that is likely to be true if the input information is. Inference is involved not only in thinking but also in perception and in motor control. When you see a three-dimensional object, a house or a horse, for instance, you have sensory information only about the part of its surface that reflects light to your retina, and perceiving it as a house or as a horse involves *inferring* from this sensory information about a surface the kind of three-dimensional object it is. When you decide to grasp, say, a mug, you use your perception of the mug and of your own bodily position in space to infer at each moment throughout the movement the best way to carry out you intention. Inferences so understood are performed not only by humans but by all species endowed with cognitive capacities. They are an essential ingredient of any cognitive system.

Even if we restrict ‘inference’—as we will do from now on—to refer to processes that have conceptual representations both as input and output (in contrast to perceptual inferences that have sensory input and to motor control inferences that have motor command output), we have grounds to assume that a vast amount of conceptual inferences are unconsciously performed all the time in human mental life. Here are some instances:

1. You hear the sound of steps growing louder and you assume that someone is coming closer
2. Joan suddenly turns and looks in a given direction and you assume that she is looking at something of possible relevance to her
3. You feel a few drops of rain falling and you assume that it is beginning to rain
4. You feel nauseous and you assume that you have eaten something bad
5. You hear the doorbell rings and you assume that there is someone at the door
6. Having heard the doorbell ring and assumed that someone is at the door, you take for granted that it probably is the person whom you were expecting at that moment
7. You are told that Bill is eight years old and that Bob is six, and you immediately realize that Bill is older than Bob.

In all these cases, there may be a reluctance to concede that some inference has taken place. From a cognitive point of view however, the fact is that some new assumption has been arrived at on the basis of previous information that warrants it. Some process must have occurred, even if rapid, automatic, and unconscious. Whatever the particular form of such a cognitive process, its function is to produce new assumptions warranted by previous ones and this is enough to make it an inferential process. In our examples, the conceptual output of a perception process (1 to 5) or of a wholly conceptual process (6 and 7) provides premises for a further inference that may be warranted inductively (1 to 6), or deductively (7). Of course, it is possible to draw such inferences in a conscious and reflective manner, but typically, the output of inferential processes of this kind is arrived at without any awareness of the process itself and comes to mind as self-evident. In other words, much inference is just something that occurs in us, at a sub-personal level, and not something we, as persons, do.

How are such spontaneous, mostly unconscious inferences achieved? One possible view is they are all drawn by a general inferential mechanism that has access to a wide base of encyclopaedic data with much general information that can be represented in conditional such as “if movement sounds are growing louder, then the source of these sounds is getting nearer”, or “if the doorbell rings, someone is at the door.” These serve as the major premise in inferences from a general proposition to a particular case where a specific assumption (e.g. “movement sounds are now growing louder,” “the doorbell is ringing”) serves as the minor premise. In other words, unconscious inference would resemble conscious reasoning with conclusions following from premises in virtue of quite general deductive or inductive warrant relationships.

Assuming that unconscious inference works like conscious reasoning except for the fact that it is unconscious raises the following puzzle: Why would we ever bother to perform in a conscious and painstaking manner what we would be quite good at doing unconsciously and almost effortlessly? Even assuming some satisfactory solution to this puzzle, we would still be left with two more substantial problems, one having to do with relevance, and the other with efficiency.

Here is the relevance problem: For any minor premise such as “the doorbell is ringing,” there is a vast array of encyclopaedic information that could provide a major premise (for instance “if the doorbell is ringing, electricity is working properly,” “if the doorbell is ringing, pressure is being exerted on the bell button,” “if the doorbell is ringing, air is vibrating,” and so on). In a given situation very few if any of these possible major premises are used to derive a conclusion, even though these conclusions would all be similarly warranted. What happens is that only contextually *relevant* inferences tend to be drawn (Van der Henst, 2006; Van der Henst, Sperber, & Politzer, 2002). A general inferential ability is not, by itself, geared to homing in on such relevant inferences.

Here is the efficiency problem: Inferences in many specific domains could be more efficient if tailored to take advantage of the specificity of the domain and to employ dedicated procedures rather than have that specificity be represented in propositional form and used by general procedures such as conditional inferences. For instance, we expect not only humans but also other animals to be able to draw inferences about the movement of looming objects from their increasing loudness. Presumably, animals do not draw these inferences by using propositional generalisations as premises in conditional inference. More plausibly, they take advantage of the regular correlation between noise and movement. This regular correlation has allowed the evolution of an ad hoc procedure that directly yields an assumption about relative nearness of a moving object from a perception of increasing or decreasing movement noises, without a general assumption about the relationship between the two being represented as a piece of encyclopaedic knowledge and used as a major premise in some form of general conditional reasoning. For humans too, it would be advantageous to have the same kind of ad hoc, automatic procedure. Similarly, understanding that someone is at the door when the doorbell rings is better left to an ad hoc kind of cognitive reflex than to general knowledge and inferential capacities. In the case of the doorbell, the 'cognitive reflex' is obviously acquired and is preceded by the acquisition of the relevant encyclopaedic knowledge whereas in the case of looming sources of sound, it might well be innate (and the relevant encyclopaedic knowledge is acquired later, if at all). What both examples suggest, in spite of this difference, is that many inferences may be more effectively carried by specialised domain-specific mental devices or 'modules' than by a general ability drawing on a single huge data base.

Assuming that the efficiency problem is, at least in part, solved by the existence of many specialised inferential modules contributes to solving the relevance problem. Presumably, those inferences that become modularised either in evolution or in cognitive development are those that are the most likely to be relevant and to be performed spontaneously (for other and subtler ways in which modularity contributes to relevance, see Sperber, 2005).

The image of spontaneous inference that emerges through these remarks (and that we have developed elsewhere: Mercier & Sperber, 2009) is not one of a single system but of a great variety of narrowly specialised modules. Several dual systems theorists have similarly argued that so-called 'system 1' is in fact a medley of diverse procedures (see of instance Stanovich, 2004). But what about system 2?

Metarepresentational inferences: intuitive and reflective

Humans have the 'metarepresentational' ability to represent representations, mental representations such as the thoughts of others and their own thoughts, and public representations such as utterances. They draw intuitive inferences about representations just as they do about other things in the world, for instance:

8. Seeing Joan open the fridge with an empty beer mug in her hand, you infer that she wants *to drink beer*.
9. From the same behavioural evidence, you also infer that Joan believes that *there is beer in the fridge*.
10. Knowing that Joan wants *to drink beer* and believes *there is beer in the fridge*, you infer that she will look for beer in the fridge.
11. You are asked whether Joan is expecting *Bill to come to the party* and knowing that she believes that *Jill is coming to the party* and that *Jill always brings Bill to parties*, you immediately answer, 'Yes!" On what basis? You infer Joan's expectation from her beliefs.
12. Asked whether she would like to go for a walk, Joan answers, shaking her head, "I am tired" and you infer her to mean that, no, *she does not want to go for a walk because she is tired*.

In 8 to 12, the italicized words represent not (or at least not directly) a state of affair but the content of a mental representation. As in cases 1 to 7, an inference takes place without, usually, being noticed as such. In 8 and 9, a mental state is inferred from observation of behaviour. In 10, an expectation of behaviour is arrived at on the basis of knowledge of mental states. In 11, a mental state is inferred from knowledge of other mental states. In 12, the content of a very specific type of mental state, a communicative intention, is inferred from verbal behaviour.

We now look at an example similar to 12, but with an interesting twist:

13. You ask Steve whether he believes that Joan would like to go for a walk, and he answers, shaking his head, "She is tired."

As Joan is in 12, Steve, in 13, is describing a state of affairs from which you can infer—and he intends you to infer—that Joan would not want to go for a walk. When Joan is herself speaking as in 12, given that people are, in such matters reliable authority on their own wants, you are likely to trust her and to believe that indeed she doesn't want to go for a walk (you may be more sceptical of her excuse, but this is beside the present point). In 13, Steve is less of an authority on Joan's wants and hence you are less likely to accept his opinion on trust. On the other hand, let us assume, you trust and believe him when he says that she is tired, something easier for him to ascertain. In saying that Joan is tired, Steve provides you with an argument for the conclusion he wants you to draw: you may yourself conclude that Joan, being tired, is unlikely to want to go for a walk.

You might have been wholly trustful and have accepted both Steve's explicit meaning—that Joan is tired— and his implicit meaning—that she does not want to go for a walk—without even paying attention to the fact that the former is an argument for believing the latter. We are, however, considering the case where, not being wholly disposed to take Steve's word for it, you pay attention to the argument. Only if you find the argument good enough in the circumstances will you then accept its implicit conclusion.

If you ponder Steve's argument, if you consider whether accepting that Joan is tired is a good reason to believe that she would not want to go for a walk, then what you are engaged in is reflective inference, you are reasoning, in a quite ordinary sense of the term. You are paying conscious attention to the relationship between argument and claim, or premises and intended conclusions. Unlike what happens in intuitive inference, you may end up accepting a conclusion *for a reason represented as such*.

Reasoning so understood involves paying attention to the relationship between claims and reasons to accept them. While accepting or rejecting the claim is done reflectively, the relationship between claim and reasons is intuitively assessed: you intuitively understand, say, that the fact that Joan is tired constitutes a good reason to believe that she wouldn't want to go for a walk. You then, or so it seems to you, decide to accept Steve's conclusion.

It could be argued that, when one consciously reasons and sees oneself as involved in a series of personal epistemic decisions, one is mistaking the visible tip of a mental iceberg for its largely invisible structure. Actually, what happens is that a series of inferences is taking place unconsciously. These inferences deliver as their output conscious representations about relationships between reasons and claims. The conscious self's grasp of these relationship is just the intuitive awareness of the output of mental work performed at a sub-personal level. The conscious self builds a narrative out of formal relationships. It sees itself as making epistemic decisions to accept or reject conclusions, when in fact these decisions have been all but made at this sub-personal level. True, one can engage into higher order reasoning, that is, reason about reasons rather than just accept them or reject them as intuitively strong or weak. You may, for instance, ponder the extent to which Joan's tiredness provides a good reason to believe she would not want to go for a walk : after all, being tired may sometimes motivate one go for a walk rather than, say, keep working. Such higher order reasoning is relatively rare, and, ultimately, cannot but be grounded in intuitions about even higher order relationships between reasons and claims. Reasoning as we consciously experience it, that is, as a series of conscious epistemic assessments and decisions, may well be, to a large extent, a cognitive illusion (just as may be practical decision, see, e.g., Soon, Brass, Heinze, & Haynes, 2008). In challenging the Cartesian sense of reasoning as an exercise of free will guided by reasons rather than compelled by causes, we challenge also the sense of reasoning as an higher form of *individual* cognition.

Here is a real-life example of reasoning in the sense intended. During the 2008 primaries of the American presidential election, the organisation Move On ran a competition for the best 30s video ad for Barack Obama. One of the winners was entitled "They Said He Was Unprepared..."[2] This is what you saw and heard:

> (*A succession of still pictures of Obama campaigning*)
> A man from Illinois was running for president.
> His opponents ridiculed him as inexperienced and woefully unprepared.
> His only government experience had been servicing the Illinois State legislature plus two years as an obscure member of Congress.
> He had never held an executive or management position of any kind.
> Yet THIS man (*now a picture Abraham Lincoln*) was elected President. Twice. And (*now pictures of Lincoln and Obama side by side*) they said he was unprepared!

Until you saw the picture of Lincoln, you assumed that the "man from Illinois" said to be unprepared was Obama, and thought you recognised arguments that were indeed being used against him at the time by his rivals for the Democratic nomination. When the picture

[2] By Josh Garrett, visible at http://www.youtube.com/watch?v=LuVNZPoVPYg.

of Lincoln appeared and you heard that “This man was elected President. Twice,” you understood that all these very arguments had been used to claim that Lincoln would not be a good president. Well, he was a great president. So these arguments against Lincoln were not good ones. By parity of reasoning, you were given to understand, the same arguments were not good ones against Obama.

Note the cognitive complexity involved in comprehending such an ad. Viewers had to correct their first impression that the target of the arguments quoted was Obama and understand that it had been Lincoln. They had to realize that what these arguments were intended to show had been refuted by Lincoln’s career. They had to focus on the fact that almost identical arguments were now used against Obama (a step made easy by their initial misidentification of the target). Finally they had to conclude that, by parity of reasoning, these arguments were flawed when levelled at Obama. Watching this ad, viewers don’t just end up with the conclusion springing unannounced to the mind as in intuitive inference that Obama’s relative unpreparedness need not stand in the way of his becoming a good president. To come to this conclusion, they have to be aware of the intermediary steps. As most real life complex arguments, this one was enthymematic, so most of these steps had to be reconstructed. Still, viewers had little difficulty doing all this in 30 seconds (as evidenced by the fact that they voted this ad the best in the competition). The almost exhilarating sense of cognitive control provided by understanding and accepting (or rejecting) such a complex argument is based, we suggest, on the efficacy of the unconscious processes involved and on the fact that the conscious self is given the grand role of undersigning their output.

It is contentious whether other animals have any metarepresentational ability, in particular the ability to infer what another animal wants or believes from observations of its behaviour. In any case, no one has ever suggested that the metarepresentational abilities of other animals, if they have any, might extend to metarepresenting not just mental representations but also public representations such as utterances or, even more implausibly, logical and evidential relationships among representations as in reasoning. Reasoning is specifically human. It is clearly linked to language (Carruthers, 2009). Reasoning takes as input and produces as output conceptual representations that typically can be consciously represented and verbalized.

Why Do Humans Reason?

Most philosophical and psychological approaches to reasoning seem to take for granted that the role or the function of reasoning is to enhance individual cognition. There is no doubt that it often does so. There also is plenty of evidence that reasoning is fallible (see Evans, 2002, in the case of deductive reasoning), and that reasoning sometimes lowers overall cognitive performance (Dijksterhuis, 2004; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006; Wilson, Dunn, Kraft, & Lisle, 1989). Moreover, reasoning is a costly mental activity. To assess its efficiency, not only benefits but also costs have to be taken into account. When this is done, it ceases to be self-evident that reasoning is just a 'Good Thing' for which there would necessarily have been selective pressure in the evolution of the species.

More specifically, cognitive mechanisms are likely to have evolved so as to be well adapted to a species' environment and to the kinds of information that environment provides (Sterelny, 2003, In press). In this respect, the human environment is unique: much of the information available to humans is provided by other humans.

Ordinarily, in a natural environment, most information is provided by the very items the information is about. Material objects for instance reflect or emit a variety of waves (e.g. sound or light) that help their identification. The information provided by many items—stones, stars, water, fire, for instance—, whether it is rich or poor, is unbiased: it is not geared to misleading organisms endowed with the cognitive capacities to exploit this information. A species' cognitive mechanisms are likely to have evolved so as better to exploit information provided by items of importance to members of that species. This is likely to have resulted, as we suggested, in the evolution of many specialised cognitive mechanisms or modules that are each adjusted to a specific part or aspect of the environment.

There are however items in the environment that have evolved so as to mislead some of the organisms that might interact with them. This is in particular the case, across species, of predators and preys, which, for complementary reasons, may gain from not being properly identified and may use, for this, various forms of camouflage or mimicry. It is also the case, within species, of potential mates that may gain from giving an exaggerated image of their qualities, and of competitors that may gain from giving an exaggerated image of their strength. In many cases, such misinformation may succeed: edible viceroy butterflies are generally mistaken for poisonous monarchs by birds who would otherwise eat them, and that's that. The conflict of interests between source and target of information may, in other cases, have led to an evolutionary arm race, with the target becoming better and better at seeing through misleading information, and the source producing in response information that is more and more misleading. In such cases, what we should expect to find, on the target's side, are quite specialised cognitive mechanisms aimed at quite specific and repetitive forms of misinformation.

Other sources of information in the environment may be neither neutral nor misleading but on the contrary helpful, as in the relationships between nurturing mothers and offspring, or among social insects. In such cases, one may find communication abilities evolving, with organisms providing honest information not only about themselves but also about their environment. Typically, the information communicated is very specific—about the location of nectar among honeybees, for instance—and the cognitive mechanisms used to exploit it are quite specialised coding and decoding mechanisms. Whereas misinformation, mimicry or camouflage for instance, works only if the targets do not have mechanisms dedicated to recognising it, cooperative information, as among social insects, works only if the recipients have dedicated mechanisms to recognise and decode it.

Human communication stands apart not only because of its incomparable richness and importance to individual cognition and to social interaction, not only because it is not a mere matter of coding and decoding (Sperber & Wilson, 1995), but also because it is routinely used both honestly to inform and dishonestly to mislead. In other terms, humans, who stand to gain immensely from the communication of others, incur a commensurate risk of being deceived and manipulated. How could, notwithstanding this risk, communication evolve into such an essential aspect of human life? To appreciate the evolutionary significance of the problem, we compare it to the well-known dilemma of cooperation.

Cooperation is another "Good Thing" that, at first blush, should be widespread in nature. In fact it is quite rare, and there is a simple evolutionary explanation for this. While co-operators stand to gain from participating honestly, that is from paying the cost and sharing in the benefits of cooperation, they stand to gain even more from cheating, that is sharing in the benefits without paying the cost. Cheating undermines cooperation and makes it evolutionarily unstable unless it is in some way prevented or circumscribed. Cheating may be controlled if co-operators recognise cheaters and deny them the benefits of cooperation. For instance, if co-operators adopt a tit-for-tat strategy, formal models have shown, then the evolutionary instability of cooperation may in principle be overcome (Axelrod, 1984).

Communication among the members of a group can be seen as a form of cooperation, and deception as a form of cheating. So, then, why not just apply models of the evolution of cooperation to the special case of communication, and have for instance an ad hoc version of the for tit-for-tat strategy: you lie to me, I lie to you, or : you lie to me, I stop believing you (see Blais, 1987, for a more sophisticated 'epistemic tit-for-tat')? Well, whereas in standard cooperation, unsanctioned cheating is always advantageous, the goals of communicators are very often better achieved by honest communication. We communicate in particular to coordinate with others, to make request from them, and, for this, honest information best serves our goal. So, if I were to lie to you or to refuse to believe you in retaliation for your having lied to me, I might not only punish you but also harm myself. More generally, to get as much benefit as possible from communication while minimizing the risk of being deceived requires a kind of 'epistemic vigilance' (Sperber et al., In press) that filters communicated information in sophisticated ways. Systematically disbelieving communicators who have been deceitful about one topic would for, instance, ignore the fact that they may nevertheless be uniquely well-informed and often reliable about other topics (for instance about themselves). Well-adjusted trust in communicated information must then take into account the character and circumstances of the communicator and the contents of communication.

Communication is so advantageous to human beings on the one hand, and makes them so vulnerable to misinformation on the other hand that there must have been, we suggest, strong and ongoing pressure for developing mechanisms of epistemic vigilance geared at constantly adjusted well-calibrated trust. Much of epistemic vigilance focuses on the communicator: whom to believe, when, and on what topic and issue. Recent experimental work shows that children develop, from the age of three, the ability to take into account evidence of the competence or benevolence of communicator in deciding whom to trust

(Harris, 2007; Mascaro & Sperber, 2009). Though it would still deserve more extensive empirical study, this ability is well in evidence in adults (Petty & Wegener, 1998).

Judging the trustworthiness of the source of information is not the only way to filter communicated information. The content of that information may itself be more or less believable, independently of its source. What might make it more or less believable is the effect that accepting it would have on the overall consistency of our beliefs. To take cases at the two extremes, believing a contradiction would introduce an inconsistency in our beliefs, and so would disbelieving a tautology, whatever their sources. Even in less extreme cases, the very content of a claim may weigh in favour of believing or disbelieving it. If a claim is entailed by what we already believe, this is a reason to believe it (and since we may not have considered this entailment of our beliefs before, it may well be novel and relevant). This however may not be a sufficient reason: Realizing that our previous beliefs entail some implausible consequence we had not thought of before may give us a reason to revise our beliefs rather than accept this consequence. If a claim contradicts what we already believe, this is a reason to reject it. It may not be a sufficient reason either. Rejecting such a claim may be missing an opportunity to appropriately revise beliefs that may have been wrong from the start or that need updating.

Believability of a content and reliability of its source may interact. If we deem trustworthy a person who makes a claim that contradicts our beliefs, then some belief revision is anyhow unavoidable: If we accept the claim, we must revise the beliefs it contradicts. If we reject the claim, we must revise our belief that the source is trustworthy.

In a nutshell, although attending to its consistency with previously held beliefs is highly relevant to filtering newly communicated information, this cannot rationally determine an automatic acceptance or rejection heuristic. Inconsistency of a claim with our beliefs on its subject-matter or with our trust in its source calls for reflection (see Thagard, 2005). This is true not just of logical inconsistency but also of a probabilistic form of inconsistency, where, in the light of what we already believe, a novel claim is just highly improbable.

Checking inconsistency may be a powerful way to help decide what new beliefs to accept or reject and what old beliefs to revise, but it is not a simple and cheap procedure. At least from an evolutionary point of view, it would not make much sense for an organism to make the effort of checking the mutual consistency of beliefs that wholly and purely result from its own perceptions and inferences.[3] Perceptual and inferential mechanisms have evolved to serve the cognitive need of the organism. Of course, these mechanisms may occasionally err, and their errors might be revealed by some form of consistency checking. However, not only would this procedure be costly, it would itself not be error-proof. Checking, on the other hand, the consistency of communicated information makes much more sense because communicators serve their own ends, which may differ from those of their audience and are often best served by misinforming the audience. We suggest that the cost of consistency checking is worth incurring only in order to filter communicated information.

Imagine, then, in the evolution of human communication, a stage where people do not argue but make factual claims that are often highly informative and relevant, but that may also be dishonest. When trust in the communicator is not sufficient to accept what is being communicated, addressees consider the contents and check both its internal consistency and its consistency with what they already believe. When they hit an inconsistency, they have to take an epistemic decision—or so it seems subjectively—and either reject the new information or else "bet on its truth" (De Sousa, 1971) and revise their beliefs.

The same people who, as addressees, use consistency checking to sift what they are told often find themselves in the position of communicator, now addressing other consistency checkers. One way to persuade one's addressees is to help them check the consistency of what one is claiming with what they believe, or even better if possible, to help them realise that it would be inconsistent with what they already believe not to accept one's claim. The communicator is better off making an honest display of the very consistency addressees are anyhow checking. This amounts to, instead of just making a claim, *giving reasons* why it should be accepted, *arguing* for it. Once communicator resort to giving reasons, they have a use for an ad hoc logical and argumentative vocabulary ("if...then", "therefore", and so on) that is of no use, on the other hand, for making plain factual claims. This vocabulary helps display, for the examination of the addressees, the reasons why they should accept claims they are unprepared accept just on trust.

[3] Note that socialised human beings, even when alone, never have perceptions and inferences that are wholly and purely their own since their human perception and inference make use of conceptual tools acquired through cultural transmission.

Reasoning can be defined as the ability to produce and evaluate reasons. It is a costly ability: it involves special metarepresentational capacities found only among humans, it needs practice to reach proficiency, and exerting it is relatively slow and effortful. Reasoning, we argue, evolved because of its contribution to the effectiveness of human communication, enhancing content-based epistemic vigilance and one's ability to persuade a vigilant audience. The reasons reasoning is primarily about are not solipsistic, they are not for private appreciation, they are arguments used, or at least rehearsed, for persuading others.

What we are proposing then is an argumentative theory of reasoning. We are not the first to do so. Others (Billig, 1996; Perelman, 1949; Perelman & Olbrechts-Tyteca, 1969; Toulmin, 1958) have maintained that reasoning is primarily about producing and evaluating arguments. They have done so mostly on introspective grounds and in a philosophical perspective. We may be more original in doing so on empirical grounds and in a naturalistic and evolutionary perspective (see also Dessalles, 2007).

Empirical evidence for the argumentative theory of reasoning

The argumentative theory of reasoning we have briefly sketched, though still too vague in many respects, has experimentally testable consequences. More specifically, claiming that reasoning is a social competence aimed at producing argument to convince others and evaluating such arguments makes it possible to engage in 'adaptive thinking': inferring structure and performance from function. The theory predicts when reasoning should be efficient, and when it should lead us astray, and how. These predictions can be pitted against relevant results from different areas of psychology – social psychology, psychology of reasoning and decision making, and developmental psychology. We briefly do so here (for a richer review, see Mercier & Sperber, 2009, In press).

The first and most straightforward prediction of the argumentative theory is that people should be good at arguing. Any evolved mechanism should be good at performing the task it evolved for – otherwise it would not have evolved in the first place. And indeed, researchers studying persuasion and attitude change have repeatedly shown that when people are interested in the conclusion of an argument, they are much more influenced by a strong argument than by a weak one (see Petty & Wegener, 1998, for a review). Participants are also able to spot argumentative fallacies and to react appropriately to them (Hahn &

Oaksford, 2007; Neuman, Weinstock, & Glasner, 2006; Rips, 2002). They can recognize the larger, macro-structure of arguments, keep track of the commitments of different speakers and correctly attribute the burden of proof – all these being skills needed to follow or take part in an argument (Bailenson & Rips, 1996; Ricco, 2003; Rips, 1998). When it comes to production, people have no problems generating arguments supporting their views or counter-arguments attacking an alternative (Kuhn, Weinstock, & Flaton, 1994; Shaw, 1996). Dunbar and Blanchette (2000, 2001) have demonstrated that people use deeper analogies when they aim at convincing someone. More generally, researchers who have looked at actual arguments and debates, even among untrained participants, are often "impressed by the coherence of the reasoning displayed" (Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993, 362).

The developmental evidence is even more striking. Nancy Stein and her colleagues have shown that children as young as three are perfectly able to engage in argumentation (Stein & Albro, 2001; Stein & Bernas, 1999; Stein & Miller, 1993). Preschoolers can even spot argumentative fallacies such as circular reasoning (Baum, Danovitch, & Keil, 2007). By contrast, standard reasoning problems have been found not to be even worth testing until relatively late in adolescence—and even then, performance tends to be abysmal (see, e.g., Barrouillet, Grosset, & Lecas, 2000).

As we mentioned earlier, the general conclusion most commonly drawn from the psychology of reasoning is that people are not very good at it. The second and more specific prediction of the argumentative theory is that people should reason much better in argumentative contexts. Reasoning should be more naturally triggered when people have to convince other people or to evaluate arguments aimed at convincing them. Moreover reasoning should be specifically adjusted for these goals and good at achieving them, just as, say, a corkscrew, being specially designed to pull out corks, is likely to be better at doing so than at performing other odd tasks it may occasionally be used for. There is much evidence to confirm this prediction. For instance, when people want to attack alternative views, they are very good at making use of *modus tollens* arguments (Pennington & Hastie, 1993). On the other hand, half of the people tested in standard reasoning tasks lacking an argumentative context fails on *modus tollens* tasks (Evans, Newstead, & Byrne, 1993).

Even more persuasive are experiments in which exactly the same tasks are used in individual and in group settings. As we mentioned, when a task has a demonstrably correct answer that most individual participants fail to give, groups generally do much better—sometimes dramatically so (Bonner, Baumann, & Dalal, 2002; Laughlin & Ellis, 1986; Moshman & Geil, 1998). This is not an effect of enhanced motivation to perform well in group situations since monetary incentives, that have, if anything, stronger motivating force, have no comparable effect (see Camerer & Hogarth, 1999, in the general case, and Jones & Sugden, 2001, for the Wason selection task). Discussion, on the other hand, is crucial for group performance (see Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006) In a discussion, participants are able both to produce good arguments and to select the best among those produced by the group. It is unsurprising therefore that learning methods that rely on peer discussions have been found to be extremely effective (see Slavin, 1995, for review), and are now being adopted at all levels of education, from elementary school up to the MIT (Rimer, 2009).

The next prediction of the argumentative theory may seem paradoxical in that it predict that human reasoning may owe part of its effectiveness to what, from a strictly epistemic point of view, should be seen as a flaw. When people are engaged in a debate, what they look for, what is useful to them, are arguments that support their views or undermine those of their interlocutor. Finding arguments for the opposite view (unless it is in order to anticipate and refute them) is counterproductive. Reasoning, as a mechanism that allows us to find arguments in such contexts, *should* therefore be biased. More specifically, it should—and does indeed—display a confirmation bias. Across a vast range of experiments, people have repeatedly shown a tendency to only look for arguments that support their views, their hypotheses (see Nickerson, 1998, for review). It has also been shown that people mostly search for new information that supports their opinions (S. M. Smith, Fabrigar, & Norris, 2008). In psychology of reasoning, the confirmation bias has been blamed for participants' failures in most tasks, including conditional reasoning tasks (Evans, 1996), hypothesis testing (Poletiek, 1996; Tweney et al., 1980) and syllogistic reasoning (Evans, Handley, Harper, & Johnson-Laird, 1999; Newstead, Handley, & Buck, 1999). The confirmation bias is not something that people can easily suppress: it is a ubiquitous feature of reasoning. Instructions putting a special emphasis on objectivity fail to diminish the bias. On the contrary, in an experiment by Lord, such instructions caused participants to reason more, but doing so with an intact bias, they provided responses that were *even more biased* (Lord, Lepper, & Preston, 1984)!

Is the confirmation bias therefore an aspect of reasoning that may be effective from a practical point of view but that makes reasoning epistemically defective? Not really. People are quite able to falsify ideas or hypotheses… w*hen they disagree with them*. When a hypothesis is presented by someone else, participants are much more likely to look for falsifying evidence (Cowley & Byrne, 2005). When, for instance, people disagree with the conditional statement to be tested in the Wason selection task, a majority is able to pick the cards that can effectively falsify the statement, thereby successfully solving the task (Dawson, Gilovich, & Regan, 2002). Similarly, when people believe that the conclusion of a syllogism is false – if it conflicts with their beliefs for instance – they are look for counterexamples, something they fail to do otherwise (Klauer, Musch, & Naumer, 2000).

Even useful biases can have excessive consequences, especially when at work in non-standard contexts. The ability to "find or make a reason for everything one has a mind to do" to use Benjamin Franklin's apt characterization (Franklin, 1799), can lead to a biased assessment of arguments when they are encountered outside of an actual discussion (Cacioppo & Petty, 1979; Edwards & Smith, 1996; Klaczynski & Lavallee, 2005). Sometimes the bias is so strong as to make people change their mind in a direction opposite to that of the argument they are presented with and that has a conclusion opposite to their own views (Lord, Ross, & Lepper, 1979; Pomerantz, Chaiken, & Tordesillas, 1995; Taber & Lodge, 2006). A similar polarization can also occur when people are reasoning on their own on some topic: finding only arguments that support their initial intuition, people end up with, if anything, a stronger view (Chaiken & Yates, 1985; Sadler & Tesser, 1973). The same mechanism leads to overconfidence in the correctness of one's answers: 'Surely it must be right, given all the supporting evidence I can think of' (Koriat, Lichtenstein, & Fischhoff, 1980). Participants also use reasoning to salvage a belief they hold dear even if it is shown to be erroneous (Guenther & Alicke, 2008). Finally, by finding handy excuses and justifications, reasoning can allow us by-pass our own moral intuitions (e.g., Bandura, 1990; Valdesolo & DeSteno, 2008). With alarming frequency reasoning on our own leads to a distortion of our beliefs, something one would not predict if the function of reasoning were to guide us towards the truth.

How should reasoning affect decision? According to the standard view, reasoning should help us take better decisions. For the argumentative theory however, reasoning is more likely to bring us towards decisions that we can justify to others, that is, decisions for which we easily find arguments. The correlation between the ease with which a decision can be justified to an audience and its rationality is at best a weak one. Often, easy justifiability may favour a worse decision. Researchers working within the *reason-based choice* framework have demonstrated that participants often choose a particular alternative because it is easy to find reasons for it (Shafir, Simonson, & Tversky, 1993). More often than not, this leads participants towards a choice that does not truly satisfy them. We would argue that the overuse of reasoning in search of justifications helps explain many errors and biases such as the sunk cost fallacy (Arkes & Ayton, 1999), the attraction and compromise effects (Simonson, 1989), the disjunction effect (Tversky & Shafir, 1992), or preference inversion (Hsee, Loewenstein, Blount, & Bazerman, 1999) (for a more exhaustive list, see Mercier & Sperber, In press). Participants tend, for instance, to choose a bigger chocolate in the shape of a cockroach over a smaller heart shaped one because it is easy to justify picking a bigger chocolate and hard to justify the 'irrational' feeling of disgust its shape may elicit—and they end up not enjoying at all the roach-shaped chocolate (Hsee, 1999)!

Some social consequences of a social competence

Reasoning, we have argued, is a specialised metarepresentational competence with a primarily social cognitive function. It is both structurally and functionally quite different from intuitive inferential mechanisms that have a primarily individual cognitive function. Collective cognitive performance may be based on the aggregation of individual intuitions or on argumentative interaction, with quite different outcomes. Individual intuitions are not aimed at collective aggregation, and when some aggregation does take place, it is typically through some quite artificial mechanism. Individual opinion on some numerical value may, for instance, be collected, and the mean computed. Provided individual opinions depart randomly from the true value, the aggregation process turns out to be remarkably efficient (see Hogarth, 1978 and Larrick & Soll, 2006, for a recent review).

When people in a group must come to some collective judgement or decision and cannot argue to do so, the group typically converges on the average of the opinions of its members

(Allport, 1924; Farnsworth & Behner, 1931). When people argue, however, the direction the group takes depends on the strength, the number and the direction of the arguments that are used (Isenberg, 1986; Vinokur, 1971; Vinokur & Burnstein, 1978). When the questions debated are relatively simple—as in many experimental settings—one argument often wins and determines the decision of the group (see for instance McGuire, Kiesler, & Siegel, 1987). The efficiency of reasoning is evidenced by the fact that when a demonstrably correct answer is defended within the group, the arguments that support this answer tend to be accepted (Bonner et al., 2002; Laughlin & Ellis, 1986; Moshman & Geil, 1998).

When argumentation and hence reasoning are at work, they shape the outcomes of group processes. In many cases, this is for the best—more information is shared, superior arguments are granted more weight. Sometimes, however, reasoning creates a *polarization* of the group (Sunstein, 2002). This mostly happens when people are forced to debate an issue on which they already agree. In this case, group members submit different arguments all supporting the same position. Other group members, agreeing with the conclusions of these arguments, do not examine them thoroughly. People thus end up with even more reasons to hold their initial view or with reasons to hold even stronger view of the same tenor. Various disasters—most famously, the Bay of Pigs (Janis, 1982)—have been blamed on this kind of process. It is important to note that in these cases, reasoning is *not* used in its normal context. There may be a discussion, but it is a forced one: people do not spontaneously argue when they agree. These results are nonetheless relevant because such situations are quite common in a modern environment. Often, groups—committees or courts for instance—charged with making a decision have to justify it, and therefore to produce arguments. When they agree on the decision in the first place, this may result in over-biased arguments.

The phenomenon of group polarization helps explain another cognitively and socially relevant feature of reasoning: its potential creativity. Intuitive mechanisms for aggregating information never lead to an answer that is outside the range of the initial answers. Reasoning often does. In many cases, this will lead a group to an answer that is better than any of those that were initially entertained (Blinder & Morgan, 2000; Glachan & Light, 1982; Laughlin, Hatch, Silver, & Boh, 2006; Michaelsen, Watson, & Black, 1989; M. K. Smith et al., 2009; Sniezek & Henry, 1989). In other cases, however, this may lead to an answer worse than any of the initial ones. In the right institutional environment, however, such excesses can themselves be turned to good. Consider, for instance, different scientific groups (labs or schools of thought) each following with utter conviction a different idea. Each group is likely to suffer from a form of polarization. When, however, there is a process of selection going on at a higher level—when for instance the ideas coming from these different groups are evaluated and tested by a larger community—then the polarization may have allowed for a much broader exploration of the space of ideas. Many will have been wrong, but hopefully some may have been even 'more right than they thought', and polarization will have allowed them to dig into new and otherwise unreachable territory.

It is common to think of science as epitomizing the power and creativity of human reasoning. Of course, it is well understood that science is a collective enterprise, but still, individual scientists are seen as contributing to it through the exercise of a faculty aimed at individually seeking the truth. Seeing reasoning as primarily a social competence aimed at persuading and at being persuaded only by good reasons suggests another way of articulating reason and science and, more generally the cognitive and the social.

A proper understanding of group performance—of 'collective wisdom' for instance—requires attending equally to cognitive and to social mechanisms.

References

Allport, F. (1924). *Social Psychology*. Boston: Houghton Mifflin.

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.

Axelrod, R. (1984). *The Evolution of Co-operation*. New York: Basic Books.

Bailenson, J. N., & Rips, L. J. (1996). Informal reasoning and burden of proof. *Applied Cognitive Psychology, 10*(7), 3-16.

Bandura, A. (1990). Selective activation and disengagement of moral control. *Journal of Social Issues, 46*(1), 27–46.

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: chronometric and developmental evidence. *Cognition, 75*(3), 237-266.

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.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning*. Hove: Erlbaum.

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

Blais, M. J. (1987). Epistemic Tit for Tat. *The Journal of Philosophy, 84*(7), 363-375.

Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition, 28*(1), 108-124.

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*.

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology*. New York: The Guilford Press.

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.

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.

Carruthers, P. (2009). An architecture for dual reasoning. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds*. New York: Oxford University Press.

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., & Yates, S. (1985). Affective-cognitive consistency and thought-induced attitude polarization. *Journal of Personality and Social Psychology, 49*(6), 1470-1481.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

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.

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.

De Sousa, R. B. (1971). How to give a piece of your mind: or, the logic of assent and belief. *Review of Metaphysics, XXV*, 52-79.

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown.

Dessalles, J.-L. (2007). *Why We Talk: The Evolutionary Origins of Language* Cambridge: Oxford University Press.

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.

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

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. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459.

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., 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., 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.

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

Franklin, B. (1799). *The Autobiography of Benjamin Franklin*.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

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.

Guenther, C. L., & Alicke, M. D. (2008). Self-enhancement and belief perseverance. *Journal of Experimental Social Psychology, 44*(3), 706-712.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review, 114*(3), 704-732.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.

Harris, P. L. (2007). Trust. *Developmental Science, 10*, 135-138.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance, 21*, 40-46.

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., 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.

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.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, UK: Cambridge University Press.

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.

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.

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

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

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.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions : Misappreciation of the averaging principle. *Management science, 52*, 111-127.

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.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*, 1231-1243.

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.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children’s vigilance towards deception. *Cognition, 112*, 367–380.

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.

Mercier, H., & Sperber, D. (In press). “Why do humans reason? Arguments for an argumentative theory”. *Behavioral and Brain Sciences*.

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.

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning, 4*(3), 231-248.

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.

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.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision-making. *Cognition, 49*, 123-163.

Perelman, C. (1949). Philosophies premieres et philosophie regressive. *Dialectica 11* 175–191.

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The New Rhetoric: A Treatise on Argumentation*. Notre Dame, IN: University of Notre Dame Press.

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., & 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.

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.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge*. New York: Oxford University Press.

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

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.

Rimer, S. (2009, January 13). A new formula for teaching introductory physics. *International Herald Tribune*.

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.

Sadler, O., & Tesser, A. (1973). Some effects of salience and time upon interpersonal hostility and attraction during social isolation. *Sociometry, 36*(1), 99-112.

Schacter, D. L. (1987). Implicit Memory: History and Current Status. *Journal of experimental psychology. Learning, memory, and cognition, 13*(3), 501-518.

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.

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

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

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

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

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

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.

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*, 543–545.

Sperber, D. (2001). In defense of massive modularity. In E. Dupoux (Ed.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 47-57). Cambridge, Massachusetts: MIT Press.

Sperber, D. (2005). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind: Structure and Contents*.

Sperber, D. (2006). Why a deep understanding of cultural evolution is incompatible with shallow psychology. In N. J. Enfield & S. Levinson (Eds.), *Roots of Human Sociality* (pp. 441-449). Oxford: Berg.

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

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (In press) “Epistemic vigilance”. *Mind & Language*.

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

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

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., & 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.

Sterelny, K. (2003). *Thought in an Hostile World*. Oxford, UK: Blackwell.

Sterelny, K. (In press). *The Fate of the Third Chimpanzee*. Cambridge, MA: MIT Press.

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

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

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

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*, 1623-1626.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

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.

Uleman, J. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology*. New York: The Guilford Press.

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*.

Van der Henst, J.-B. (2006). Relevance effects in reasoning. *Mind & Society, 5*(2), 229-245.

Van der Henst, J.-B., Sperber, D., & Politzer, G. (2002). When is a conclusion worth deriving? A relevance-based analysis of indeterminate relational problems. *Thinking and Reasoning, 8*, 1-20.

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. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology: I* (pp. 106–137). Harmandsworth, England: Penguin.

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., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101-126.