1

THE SOCIAL ORIGINS OF FOLK EPISTEMOLOGY

Abstract

Because reasoning allows us to justify our beliefs and evaluate these justifications it is central to folk epistemology. Following Sperber, and contrary to classical views, it will be argued that reasoning evolved not to complement individual cognition but as an argumentative device. This hypothesis is more consistent with the prevalence of the confirmation and disconfirmation biases. It will be suggested that these biases render the individual use of reasoning hazardous, but that when reasoning is used in its natural, argumentative, context they can represent a smart way to divide labor without loosing epistemic value.

1 Introduction

Folk epistemology involves a set of cognitive mechanisms that allows us to evaluate beliefs, inferences, utterances or other products of the mind. Some of these abilities are clearly social in nature. For instance, we routinely evaluate the competence and benevolence of the people we interact with (Fiske, Cuddy, & Glick, 2007), and we use this information in order to accept – or reject – what they tell us (see Petty & Wegener, 1 998, for adults and Koenig & Harris, 2005, for children). Reasoning also plays a crucial role in folk epistemology, by understanding and weighing arguments that are offered to us in support of a given conclusion, thereby allowing us to decide whether we should accept this conclusion or not. This function of reasoning, however, is usually overshadowed by the benefits it purportedly offers to individual cognition, so well described by Descartes: “The long

chains of simple and easy reasonings by means of which geometers are accustomed to reach the conclusions of their most difficult demonstrations, had led me to imagine that all things, to the knowledge of which man is competent, are mutually connected in the same way, and that there is nothing so far removed from us as to be beyond our reach, or so hidden that we cannot discover it, provided only we abstain from accepting the false for the true, and always preserve in our thoughts the order necessary for the deduction of one truth from another". In the recent literature, dual process models of reasoning claim that analytic reasoning evolved in order to allow us to follow normative guidelines and thus correct the mistakes made by our more sloppy intuitive[1] mechanisms (see, e.g. Evans & Over, 1996). Contrary to this claim, Sperber has proposed that reasoning evolved for *argumentative* purposes: so that we can convince others and evaluate their arguments when they try to convince us (Mercier & Sperber, 2009; Sperber, 2000, 2001, see also Dessalles, 2007). In this paper, we will offer a somewhat more precise evolutionary story, provide some support for it and look at some implications regarding the effects and efficiency of reasoning.

2 Classical explanations for the evolution of reasoning

It has now become fashionable to at least hint at an evolutionary rationale for the mental abilities one is studying. In the field of reasoning, some do so with reluctance – considering that evolutionary stories are mostly post-hoc and of no great explanatory value (Evans, 2006) – while others stress the importance of evolutionary thinking and even find in it the inspiration for their theories (Stanovich, 2004). Before we start reviewing what the different theories of reasoning have to say about its evolution, we must make a semantic point. Nowadays, the most influential theories of reasoning are couched in terms of *dual process*: they claim that we have two kinds of reasoning mechanisms. The first system is variously known as associative, experiential, heuristic or, more simply, system 1. It is thought to be implicit (or unconscious), automatic, fast, and prone to mistakes. The second system – analytic or system 2 – would be based on rules allowing us to follow normative guidelines and correct the wanderings of system 1. For instance, in a classical – and therefore tricky – task of reasoning and decision making, such as the Linda problem (Tversky & Kahneman, 1983), the Wason selection task (Wason, 1966) or the bat and ball (Frederick, 2005), people

[1] What is called here, following Sperber, intuitive mechanisms is also known as system 1 processes, or heuristic system. I prefer the term intuitive because: (i) it is more explicit than 'system 1', (ii) reasoning also has a very heuristic nature (in fact more so than most intuitive processes) and, (iii) it corresponds quite well with the naïve notion of intuition – something we feel is right without knowing the reasons.

often experience a conflict between an intuitive (but wrong) answer suggested by system 1 and the normative answer that the system 2 came up with or accepted. It can be argued however that using the term 'reasoning' to describe system 1 processes is somewhat misleading. These processes are mostly unconscious and involve more a sense of 'intuition' (see Hammond, 1996). They are responsible for the first, immediate impression you have of a person for instance. This hardly corresponds to the everyday use of 'reasoning', which typically involves concentration and awareness. We will therefore use the word 'reasoning' to refer to system 2 reasoning only, and we will use 'intuitive inferences' for system 1 processes (see Sperber, 1997).

A more detailed account – and criticism – of the evolutionary arguments advocated in dual process theories can be found in Mercier and Sperber (2009), so we will keep it extremely short here. Steven Sloman made two suggestions regarding the function of reasoning (Sloman, 1996). It could either complement the rest of cognition by allowing the focus on relevant features of a given object, or it could help repress impulses and delay gratification. These abilities however are quite widespread in the animal kingdom and it is hard to see how they could predict the specific features of reasoning. Evans and Over suggested that reasoning evolved in order to "deal with novelty and anticipate the future" (Evans & Over, 1996, p.154). Again, all animals have to some extent to deal with novelty and anticipate the future. This would be the more general function of learning and (prospective) memory, allowing us to draw lessons from the past in order to behave more appropriately in the future. But it is Stanovich who offers the most in depth evolutionary account. Specifically, he states that reasoning is "where the genes gave up direct control and instead said (metaphorically, by the types of phenotypic effect that they created) 'things will be changing too fast out there, brain, for us to tell you exactly what to do – you just go ahead and do what you think is best given the general goals (survival, sexual reproduction) that we have inserted.'" (Stanovich, 2004). For him, intuitive inferences are given by genes in the form of rules such as 'when in context Y, do A'; such rules favor the genes over the individual. Reasoning allows us to inhibit these processes and decontextualize our thinking, thus granting us more flexibility and delivering us from the grasp of our genes. A similar criticism can be applied to this theory: this function is hardly specific of reasoning – such ability to decontextualize (or, more accurately, to take context into account in a finer grained fashion) is sometimes attributed to the frontal cortex generally and can be used to explain the behavior of rats for instance (Prescott, Redgrave, & Gurney, 1999).

It would thus seem that all of these proposals fail to give an evolutionary rationale for the specific properties of reasoning. This may stem from

the fact that they all see reasoning as a tool correcting the shortcomings of other cognitive mechanisms and enhancing individual cognition. The general argument against this view can be articulated as follows: (i) The whole of cognitive activity (in most animals) or at least the vast majority of it (in humans) goes on without reasoning. (ii) There is no reason to expect that evolution has been able to achieve with reasoning a cognitive procedure that would be so efficient as to be able to correct for the mistakes of numerous other mechanisms—especially given the numerous mistakes it can lead to, as will be detailed in the following (see also Mercier & Sperber, submitted). (iii) Reasoning is costly. If we grant these points, then the advantages reasoning offers as a tool to enhance individual cognition are not clear. This argument is, admittedly, not a proof. However, it shows that we should not accept as obvious the classical story about the individual origin of reasoning. Ultimately, the argument will be won by the theory that fits better with the data we have on reasoning. What follows is an evolutionary story that could more plausibly have led to the evolution of reasoning mechanisms, and a short review of the data supporting it.

3 The dangers of communication

Communication is dangerous. You can be lied to, deceived, manipulated. These dangers are well known to biologists who study the evolution of communication. They point out that for communication to evolve it has to be mostly honest: if senders don't benefit from it, they stop sending, and if receivers don't benefit, they stop receiving (Krebs & Dawkins, 1984). Maintaining honesty is hard though because each individual sender would benefit more if she could manipulate receivers to her advantage. One way to maintain honesty is for communication to be costly: if you have to incur a cost in order to be able to communicate, and if only some individuals are able to incur this cost, then communicating reliably signals your ability to pay this cost (Zahavi & Zahavi, 1997).

Human communication however is mostly 'cheap talk': speaking doesn't cost us much in terms of energy for instance. Accordingly, they have evolved a suite of mechanisms to deal with this issue (Sperber et al., In press) Such a solution to enforce honesty is for receivers to calibrate their trust appropriately and to stop responding to senders who have sent dishonest signals. This is called "skeptical responding" and some monkeys have been shown to make a smart use of this strategy (Cheney & Seyfarth, 1990; Gouzoules, Gouzoules, & Miller, 1996). As we mentioned in the introduction, humans too make a wide use of this strategy: we calibrate our trust following criteria of competence and benevolence and react accordingly to the information we are communicated. However, humans also use

other strategies. Most importantly, they test the coherence of what they are told against what they already know.

4 Coherence checking

Maintaining a coherent behavior has been important at least since the apparition of bilateral symmetry (Koopowitz & Keenan, 1982; Prescott, 2007). Several structures are important in this respect. Some deal with action selection and making sure that the behavior remains coherent through time (most famously the basal ganglia, see Redgrave, Prescott, & Gurney, 1999). Others are thought to be in charge of detecting conflicts within the system (such as the anterior cingulate, Botvinick, Braver, Barch, Carter, & Cohen, 2001). Some forms of cognitive dissonance, such as that found in other species, could be explained by the necessary presence of such regulatory mechanisms in any complex cognitive system (see Egan, Santos, & Bloom, 2007).

This sensitivity to dissonant information could be used as a tool to filter incoming information. If someone tells me something that clashes with my beliefs, then I can use this as ground for rejections, or at least further examination. If the mechanisms that deal with incoherent information can be used, they also have to be modified in order to fulfill this sifting function. First, the preferences for new versus old information must be reversed. Generally in case of conflict (such as between perception and memory) old information should be forgotten, or at least updated, in view of new information. In the case of communication, it is more cautious to do the opposite: if you think there is food in A and someone tells you there is none, you might want to stick with your initial belief. Moreover, such conflicts can be used to modify the trust we grant different speakers. Lowering the trust granted to a speaker who says things that are inconsistent with our beliefs is a form of punishment – it in part deprives her of the benefits of communication – and can therefore contribute to maintaining the honesty of communication. Humans seem to follow both of these rules. They have in intrinsic bias towards their own opinion compared with that of others (Krueger, 2003). And when someone tells them things that are incoherent with what they think, they can call her a liar (DePaulo et al., 2003) with all that ensues, or at least drastically diminish the trust they grant her (Slovic, 1993; Yaniv & Kleinberger, 2000).

5 Bypassing coherence checking: the evolution of argumentation

Faced with such sophisticated receivers senders can try to become even more sophisticated and an arms race ensues. Receivers use coherence with their beliefs as a mean to evaluate communicated information. However they cannot check the coherence with all of their beliefs: they will only use the ones that are the most relevant at the time and given what has been communicated. For example a Frenchman (Paul) who is not a huge soccer fan hears that Brazil won the world cup in 1964. There is no clash between that and what is immediately brought to mind – it would be different if the year was 400 BCE or the team that of Laos – so he accepts this information. However, it is very likely that somewhere in his belief base are the two following beliefs: France won in 1998 and soccer world cups occur every four years. If he had taken these two beliefs into account, he would have been able to reject the communicated information. Now imagine that this person then repeats this information to someone who is more knowledgeable soccerwise (Valérie). For her this information will immediately call to mind the fact that Brazil won in 1962, and in conjunction with the premise that world cups occur only every four years, lead her to reject the communicated information. Valérie could then try to convince Paul that Brazil could not have won in 1964. She could just say "that's impossible because Brazil won in 1962", but Paul – sharing the stereotype that guys are more into soccer – might not take Valérie's word for it. She will have to find premises that Paul will accept and that will convince him that she's right. The best might be to remind him that France won in 1998, and that since world cups occur every four year, there was no world cup in 1964. The question is: how can she do that? How can she find the relevant premises among the huge quantity of soccer related information that any soccer fan possesses? We will argue that this asks for a specialized mechanism[2] that looks for relevant premises – and that this is the primary function of reasoning.

To understand that we need a specialized mechanism for this task, it must first be shown that other mechanisms that are already present cannot accomplish this function. What mechanisms could be useful in that context? First, individuals must be able to communicate effectively. They must be able to find the means so that others understand what they want to convey. However, this cannot be enough to find premises. Going back to Valérie, her aim is to communicate that "Brazil didn't win in 1964". Her communicative abilities will allow her to find the right words so that she can be understood by Paul, but nothing else. These mechanisms deal with being understood, not being convincing. Another mechanism that could be used is

[2] By specialized mechanism I mean 'module', but not so much as the word is often used in the cognitive sciences (following Fodor), but in the more general sense of a mechanism whose inner workings are mostly independent from other mechanisms.

the ability to determine whether the other person is going to accept what we say. As we mentioned, people who say things that are not coherent with their audience's beliefs tend to suffer from a decrease in trust (try urging people to vote for Le Pen at a socialist meeting). To protect themselves from this risk, individuals must be able to "think twice before saying something" or, more precisely, to filter out communications that would be inadequate. But this again cannot be sufficient. This mechanism could warn Valérie that she cannot bluntly say that "Brazil won in 1962" because she might not be believed and Paul could not only think that she knows nothing about soccer, but also that she is mildly annoying. But this mechanism will not tell her what the premises are that would allow her to persuade Paul: there are too many things that Paul would be willing to accept but that won't do the trick as premises.

Lastly, general phenomena of relevance and availability will play a role. When the topic is soccer, soccer related information will tend to be more easily available. However in this case too there is too much information available to pick up from. Moreover, in the normal course of cognition what is relevant is to draw new inferences, not to come back to why we think what we think. The facts that France won in 1998 and that world cups occur every four years are highly relevant as premises to the conclusion that Brazil didn't win in 1964, but they do not follow from it at all. All of these mechanisms can serve to restrict the range of beliefs that will be considered as potential premises: they must be communicable, acceptable by the other, and relevant to the general topic at hand. But these mechanisms cannot determine what specific beliefs will be relevant *as* premises.

Learning could have played an important role during the evolution of reasoning, because the problem of finding premises might not have been so stark for our ancestors. If we assume that their communicative system was more limited in content than ours, and given that only communicable beliefs can be taken as premises, it follows that the search space for premises might have been considerably smaller. In these conditions, it is possible that the mechanisms mentioned above might have allowed individuals to find appropriate premises. The receivers would be an important source of feedback: if the premises are appropriate, they will be convinced, if they are not, they won't. This allows for an immediate feedback that has two nice properties. The first is that failing has a cost. When we use an inappropriate premise we will have laboured in vain (the other person will not accept the intended conclusion) and – more importantly – our reputation might suffer. This will happen because the conclusion is something that the receiver will think is incoherent with her beliefs – that's why we were looking for a premise in the first place. The second is that this cost need not be very high. It makes sense that most persuasion attempts – especially at this stage – will

involve conclusions that are not strongly incoherent with the receiver's beliefs, both because of the higher price for failure (the other might think of us as a fool or a crook) and because of the difficulty of finding convincing premises in this case. So if the persuasion attempt fails, given that the conclusion is not highly incoherent with the receiver's belief, she will probably lower the trust she grants us only minimally. This is quite different from, say, a snake detector mechanism for which the feedback is less lenient, and allows for a more comfortable and extended learning phase. But even if the cost is low, it is still there, and it will favour individuals who learn faster. So we have both a possibility for the skill to be learned, and a selection pressure for this learning to be made easier – to be more 'prewired' so to speak.

It may be worth belabouring a point that might seem to go against some ideas developed earlier by Sperber (2001, for instance). The claim made here is that reasoning evolved firstly in order to bypass the coherence checking mechanisms of other individuals – in other words, to persuade an audience. Earlier, it was its role in detecting deception attempts that was stressed – its filtering role as a mechanism of epistemic vigilance. So, what role was the more important during the evolution of reasoning? Two arguments can be made to support the view that persuasion was primary. The first is that if no one is producing arguments, then there are no arguments to examine. The second is that at any given point of the evolution of communication, it has to be stable. This means that the receivers have to benefit from it overall. In order not to be too easily taken advantage of by senders, they should then tend to err on the side of cautiousness: they will be rejecting too much information rather than not enough. That puts pressure on senders to find ways to have more information accepted. This also means that new mechanisms of epistemic vigilance will necessarily allow receivers to accept *more*, not less, information. The contrary would mean that they were accepting too much information for their own good before, making communication unstable. This opposition is not necessarily of great import though: the mechanisms used to produce arguments and to evaluate them are profoundly similar. In both cases, propositions have to be evaluated in order to see if they are fit as premises for a given conclusion – either because a sender is looking for a premise to convey a conclusion, or because a receiver is trying to evaluate an argument. So even though I favour the idea that reasoning evolved first as a persuasion device, the fact is that everyone uses it both ways, and that given the overlap in functional requirements, this distinction between persuading and filtering should not play an important role.

6 How do we recognize good premises?

We have said that a mechanism that finds good premises is necessary to account for the way we argue. We have suggested that its evolution might have been easier among our ancestors and that a good feedback mechanism allowing for learning but also favouring some innateness was present. But this doesn't tell us what the actual cues are that this mechanism can use in order to find good premises. For instance, being attacked by a predator or not is a good feedback mechanism for a predator detection mechanism. But for this mechanism to work, cues must exist that reliably differentiate a predator from a non predator (such as having both eyes facing us). Likewise, cues must exist that differentiate a good premise from a poor one. It is fair to say that we don't know what these cues are (but see Mercier, submitted, for a tentative answer). But it is important to point that this doesn't mean that there is no answer. Psychologists are only beginning to understand what cues are taken into account when we evaluate if someone would be a good friend or a good mate, if some food is edible, if an animal is dangerous and, in the case of purely learnt skills, how experts in any given domain manage to accomplish their feats. We can use evaluating potential mates as an example. When we meet someone, a host of cues can be used in determining if that person would be a good mate. Most will only play a limited role, but some will be of paramount importance. Likewise, in the case of premises lots of different cues can be taken into account, some being more decisive than others. In this case however, establishing what these cues are is much harder because we are dealing with objects that we cannot directly observe or easily manipulate – namely representations. It should therefore come as no surprise that we still don't know how we recognize good premises.

Once this ability to find premises is available, it is possible to use it in order to evaluate other people's arguments: in this case as well, people have to determine whether a given representation is a good premise for a given conclusion—the difference being that in this case, someone else has given them the premise and the conclusion. When we are confronted with a premise and a conclusion, we can use reasoning to determine if the premise is a good one. This ability that allows us to find good premises and to evaluate the strength of the link between a premise and a conclusion can properly be called reasoning.

7 Consequences for individual reasoning and decision-making

The first consequence of this view of reasoning is that we should observe some prevalent biases (see Mercier & Sperber, submitted, for an exhaustive empirical argument). When reasoning is used not in order to evaluate an argument but before making a decision or stating something, its function is not to help make a better decision (including statements), it is to find support for that decision, to justify it. This can have two broad types of consequences. If, before reasoning, we are highly confident in the decision or the statement then reasoning will tend to bring only support for this decision, and it should bolster our confidence. If, on the other hand, our confidence is not very high then reasoning can influence our decision – not necessarily towards one that is better, but towards one that is easier to justify. In support of this, a quick review of the relevant literature will show (i) that confirmation and disconfirmation biases are extremely prevalent in reasoning, (ii) that reasoning doesn't necessarily help make better decisions and (iii) that reasoning tends to make us choose options that are easier to justify.

Confirmation bias consists in « seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand » (Nickerson, 1998, p. 175). Disconfirmation bias is the other face of the same coin: it appears when we are overly critical of an opinion that doesn't fit our views. As a result of these two biases people have a strong "myside bias", meaning that they will find arguments for their side of an issue much more easily than for the other side. These three biases have been found ever and again, both in the laboratory and in natural settings, and there is no space here to review these works (Edwards & Smith, 1996; Nickerson, 1998; Stanovich & West, 2007). Two points can be emphasized though. The first is that these biases are not only very prevalent, but are also very robust. Instructions to reason objectively usually have no or very limited effect (see for instance Frantz & Janoff-Bulman, 2000; Lord, Lepper, & Preston, 1984; Poletiek, 1996). These biases do not disappear with higher education or even with scientific training (Mahoney & DeMonbreun, 1977; Toplak & Stanovich, 2003). They are found in people with high 'cognitive ability' and an open mind (Stanovich & West, 2007). The prevalence and robustness of these biases render them hard to explain in terms of cognitive limitations or random wanderings of the reason. These results fit much better with the view of reasoning as an argumentative device that seeks to find support for our views: in this case, these biases are built in the way reasoning works, and they are therefore expected to be prevalent and robust.

Another point worth stressing is that experiments that have managed to attenuate or get rid of these biases did so by changing the participants' perception of the belief that is being reasoned upon. For instance, participants tend to be lenient with studies that support their views, and overly critical of studies that contradict them (Ditto & Lopez, 1992; Edwards & Smith, 1996;

Lord, Ross, & Lepper, 1979). Instructions to be objective have no or limited effect, but if the participants are led to believe that the conclusions were in fact the opposite of what they were thought to be, then they become perfectly able to criticize the study they had considered with leniency before, and vice versa (Lord, Lepper, & Preston, 1984). Likewise, participants who seemed to be unable to use a falsifying strategy became quite apt at doing so when the hypothesis to be tested came not from them but from someone else (Cowley & Byrne, 2005).

Finally it seems that these biases have the expected effect on the beliefs we reason upon: they bolster our confidence. The confirmation bias is probably one of the explanations for overconfidence (Koriat, Lichtenstein, & Fischhoff, 1980, see also Wilson & LaFleur, 1995). When we are confident in our views, expecting to have to justify them tends to strengthen them (Lambert, Cronen, Chasteen, & Lickel, 1996). More generally, when we have to defend a decision after having made it, we tend to be more confident that it was a good one (Lerner & Tetlock, 1999). So it would seem that when we feel strongly about something, reasoning is not going to change our view – quite the contrary.

It is obvious however that in some cases reasoning will change our views on an issue, or towards a decision to be made. This should happen mostly when we are not highly confident in an option to start with, when we have no strong intuitions. When this happens however, there is still no guarantee that reasoning will help us make better decisions or have epistemically superior beliefs. In fact there even is a flurry of evidence showing that for a wide range of cases, reasoning will actually be a hindrance. When they have to think about reasons for their choices, people pick posters that will satisfy them less (Wilson et al., 1993), jellies that are judged to be less good by experts (Wilson & Schooler, 1991), apartments, roommates, cars and furniture that are less optimal (Dijksterhuis, 2004; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006), and they are less good at predicting the outcome of basketball games (Halberstadt & Levine, 1999).

Similar effects have been observed far from social psychology. In the field of implicit learning, it has been shown that for some tasks (in this case learning an artificial grammar), trying to think explicitly about the learning process disturbs it (Reber, 1993). Likewise, the oft cited failures to take base rates into account or to deal with risk and uncertainty (see Kahneman, Slovic, & Tversky, 1982) seem to be a result of an over reliance on explicit processes: when the data acquisition is implicit or when the questions are asked implicitly (or both), performance improves dramatically – reaching optimality in some contexts (Maloney, Trommershäuser, & Landy, 2007; Spellman, 1993). So the standard assumption of most economists and political scientists – but also of most psychologists of reasoning – that reason-

ing leads to better choices clearly does not hold across the board. This could be a simple disruptive effect of reasoning – it could be for instance that reasoning uses resources and renders them unavailable for other processes. This may be the case in some instances, but there is evidence that reasoning creates a much more systematic bias, a bias towards choice that are easy to justify.

The idea that our choices are guided by the reasons we can provide for them has created a successful research program around "reason-based choice" (Shafir, Simonson, & Tversky, 1993). Based on ideas from the heuristic and biases school (see for instance Slovic, 1975), it was first clearly articulated by Simonson. In a 1989 paper, he showed that the attraction effect (the fact that the introduction of an inferior alternative can modify a choice) and the compromise effect (the fact that a middle option in a set will tend to be favoured) can both be adequately explained by the use of reasons. Consistent with this interpretation, the choices based on reason were more common when participants had to justify their answers, and the justifications offered for these choices were more elaborate. Since then the "reason-based choice" framework has been used to explain many failures to follow normative guidelines (see for instance Arkes & Ayton, 1999, Shafir, Simonson, & Tversky, 1993). Basing their choices on reason seem to often lead people away from the normative answer, but it should allow them to be more persuasive if they had to convince someone that they made the right decision. Direct evidence that this is the case is hard to come by, but there is some indirect evidence. For instance, because people from different cultures will vary in the reasons they will use, having to justify their choices can make people from different cultures make different choices, even if their choices were similar when they didn't have to justify them (Briley, Morris, & Simonson, 2000). It seems very plausible that because the justifications fit so well with the culture, they would render the choice more convincing – in the same culture obviously. Also, in some cases at least, the choices that are dictated by reasons are dominant in group settings. This is the case of sunk costs for example: groups tend to commit the fallacy even more than individuals (Whyte, 1993).

So reasoning, when done individually, doesn't necessarily yield very impressive results. Does that mean that we should go with our guts all the time? Of course not, for at least two reasons. The first is that in some cases our guts can be terribly wrong. When we are confronted with a problem that we did not evolve to solve, and that we did not acquire an expertise about, then there is no reason our intuitions should be particularly good. Moreover, when the (even imaginary) audience in front of which we expect to justify ourselves is fair-minded and knowledgeable, then reasoning can indeed be a good idea. It has been shown that expecting to be accountable

to such an audience often improves performances (Lerner & Tetlock, 1999). We will now turn to another, more natural, way to improve the outcome of reasoning.

8 Consequences for group reasoning and decision making

The results reviewed above could lead us to think that reasoning is much less efficient than it actually is. This is because in these experiments reasoning was used in the confine of participants' minds, and if the evolutionary story we told is right, then this is not the normal context for reasoning to occur. If reasoning is a tool for persuasion and evaluation, it should be used when we are arguing with someone else. To see what effects these contexts have on reasoning, we will turn to the analysis of group reasoning and decision making.

When used individually reasoning is mostly a tool of (anticipated) persuasion. In group contexts however the evaluative function of reasoning will be used too. This allows people to differentially accept good and bad arguments, to filter incoming information. And they are able to do that: in the right conditions (when they are involved in the task, not distracted, etc.) people are more influenced by strong than by weak arguments (Petty & Wegener, 1998). As a result, groups too should function as filters. Assuming each member has an opinion before the discussion starts, then their reasoning should be oriented towards trying to convince others that they are right. Everyone hears out the arguments of the others and gives more weight to good arguments. The most interesting tasks in this respect are those that fulfil the following conditions: (i) they have a demonstrably correct answer that (ii) can be understood by all members and (iii) that is arrived at by at least one member (Laughlin & Ellis, 1986). This is true of some mathematical problems (Laughlin & Ellis, 1986; Stasson, Kameda, Parks, Zimmerman, & Davis, 1991), logical problems (Maciejovsky & Budescu, 2007; Moshman & Geil, 1998), 'eureka' problems (Laughlin, Kerr, Davis, Halff, & Marciniak, 1975), or problems drawn from the mastermind game (Bonner, Baumann, & Dalal, 2002). For these tasks, the social decision schemes (Davis, 1973), 'truth wins' applies: if at least one participant has the good answer, this answer will be that of the group. This means that groups will then perform as their best member, leading to an increase in performance that can be dramatic (see Moshman & Geil, 1998 for instance).

It should be stressed that the phenomenon at play here is not a simple transmission of information. It's not as if one individual was recognized, for some reason, as being the smartest or the more competent and that the others listen to her. For group processes to increase performances, conflict

is crucial (see Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006 and numerous references within). Moreover, the filtering function of groups can discriminate in a fine grained fashion and pick up only the good pieces in an individual's opinion. Take the Wason selection task for instance. It is a classical reasoning problem for which the truth wins scheme applies (Moshman & Geil, 1998). But it has been observed that a group can get to the correct answer even if no individual member had it to start with. It is enough that the pieces of the correct answer are distributed – with their correct justification – among them. In these cases, groups outperform their best individual members. This is known as the "assembly bonus effect" in social psychology (Laughlin, Bonner, & Miner, 2002; Michaelsen, Watson, & Black, 1989; Sniezek & Henry, 1989; Tindale & Sheffey, 2002) and as the "two wrongs make a right" in developmental psychology (Glachan & Light, 1982; Schwarz, Neuman, & Biezuner, 2000).

So people are able to choose positions that are better supported than their alternatives. The problem is that sometimes well supported positions are wrong. This is the case of the representativeness heuristic and the conjunction error. In the classic demonstration of the use of this heuristic a description of a person very representative of a feminist is given, and participants are asked to rank the chances that she is, among other things, a bank teller, and both a feminist and a bank teller (Tversky & Kahneman, 1983). Given that 'feminist bank teller' fits more with the description, participants tend to ascribe to it a higher probability than to 'bank teller', thus violating the axiom of probability theory stating that $P(A) \geq P(A\&B)$ and committing the conjunction fallacy. The representativeness heuristic can easily yield what could be considered as a good reason for the answer: "look as the description, she's very likely to be a feminist", and it has been observed that even when only one group member supports her answer with this heuristic, she will tend to win the group's opinion[3]. When a wrong but well supported answer is given by a majority of the members, this can strengthen the bias – much in the same way as when we are confident in our own answer, reasoning upon it will tend to bolster our views. This seems to apply for sunk costs (Whyte, 1993). Given that most people commit the fallacy, and that they seem to be doing so because of the reliance on a rea-

[3] That is, provided that the other participants propose other non normative answers. The case in which one of the participants has the normatively correct answer and the good explanation for it hasn't been tested, but given the results mentioned above and the fact that it is quite easy to demonstrate the correct answer, she should win the argument.

son, it is only to be expected that groups fare even less well than individuals[4].

9 How to make the best of a biased world

A minimal way to look at the way groups work is to consider them as aggregating, levelling mechanisms. Each member comes up with an answer, and provided that the distribution is more or less random around the correct choice, a simple rule like majority wins will yield good results in a broad range of cases (Hastie & Kameda, 2005). In this light, groups are a very simple debiasing mechanism: we are all biased towards our views, and, through averaging, groups compensate for these biases. So we might think that groups are an effective way to use reasoning because it gets rid of the biases that plague it. But doing so cancels all the added value reasoning can bring: group members might as well give their answers right away without any reasoning and justification and let the group average. Another way to look at groups is to see them as tools that allow us to make the best of reasoning.

To understand this we can compare the case of an unbiased reasoning ability with what happens in groups. What would an unbiased reasoning ability look like? If such a mechanism had to evaluate a belief for instance, it would alternatively find arguments for and against it and impartially weight these arguments. Such a mechanism would not be foolproof – we don't have infinite knowledge or infinite time and energy available – but it should generally better the epistemic status of the belief without overly tipping the scale in favour of beliefs we had a soft spot for to start with. In a way, this is exactly what groups can do – and they can even do it better. Imagine two people reasoning independently over a choice between two options. The first person favours option A, and will think of arguments confirming this view, thus inadequately bolstering her belief that she is right. Same for the other person and option B. But if these two people argue together, what should happen is similar to what would have happened if any one of them had used 'unbiased' reasoning. They are going to suggest arguments that will then be compared on the basis of their relative strength. But in this case, instead of having to find and evaluate arguments for *both* sides, each of them only has to find arguments for her side and evaluate argument for the other, thus cutting back the effort by a factor two. In that case, the very biases that are usually seen as plaguing reasoning become an

[4] Again, this is assuming that no participant is aware of the normative answer and the normative explanation. When this is the case, reasoning decreases the sunk costs fallacy (Simonson & Nye, 1992).

efficient way to divide labour. It is thus not surprising that some of the more efficient forecasting methods require people to make a forecast, justify their decisions and share this with the other forecasters in an anonymous way. The same operation is then repeated over several rounds. The Delphi method for instance (a variation on this theme) has been shown to be extremely effective, having advantages over other group settings, statistical models and prediction markets (Rowe & Wright, 1999)[5]. On the other hand, the more conditions depart from a somewhat idealized discussion context—for instance by large asymmetries in power, the introduction of a larger audience that participants may try to target instead of aiming at their interlocutor, or forced debate between people who agree on the issue at hand—the more likely it is that argumentation is going to lead people astray.

10 Conclusion

This paper opened with a general argument against the view that reasoning is mostly a prop of individual cognition. An alternative story, in which reasoning evolved as an argumentative device, was suggested. Even if one were to buy this argument, it would then be tempting to think that reasoning could have been co-opted for more general purposes: reasoning would have evolved first for argumentation but, being useful for so many other things, its feature would have changed to accommodate these new functions. Then the biases that were mentioned earlier – the confirmation bias and the disconfirmation bias – wouldn't have a reason to exist anymore – they would just be a hindrance. It is not obvious that an a priori argument can be made against the view that reasoning could have been co-opted for many different purposes besides argumentation. However, it seems that the features of reasoning that we can observe now fit much better the simple story that argumentation has always been the main driver of its evolution. People who think that reasoning subsequently evolved to fulfill a whole array of functions should now show either that the biases and other features of reasoning that fit only with the argumentative function are actually not present – and that would mean reinterpreting a considerable body of research, only hinted at here – or that these features acted as constraints that could not be overwhelmed by selection pressures.

---

[5] The benefits of information pooling have been well demonstrated in other animals, who seem to be using a form of the Condorcet Jury Theorem (Conradt & Roper, 2003, 2005; List, 2004). However, debate has features that render it quite different from other forms of opinion aggregation (such as voting – as in the CJT – or bargaining). There is no space here to discuss these issues further, but see Elster (1998) for an opinion that would be closed to the one that would be defended here.

As a final word, I cannot help but to mention how sad reasoning would be if it was a purely individual endeavour: "for man in the preservation of his being and the enjoyment of the rational life there is nothing more useful than his fellow-man who is led by reason." (Spinoza, 1677, Ethics IV app. IX).

Acknowledgments. Dan Sperber's work and his support have been instrumental in the writing of this article. I also wish to thank two anonymous reviewers and an editor for their comments and suggestions.

References

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.

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.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. 2001. Conflict monitoring and cognitive control. *Psychological Review,* 108(3), 624-652.

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.

Cheney, D. L., & Seyfarth, R. M. 1990. *How Monkeys See the World*. Chicago: Chicago University Press.

Conradt, L., & Roper, T. J. (2003). Group decision-making in animals. *Nature, 421*(6919), 155-158.

Conradt, L., & Roper, T. J. (2005). Consensus decision making in animals. *Trends in Ecology & Evolution, 20*(8), 449-456.

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.

Davis, J. H. 1973. Group decisions and social interactions: A theory of social decision schemes. *Psychological Review,* 80, 97-125.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. 2003. Cues to deception. *Psychol Bull,* 129(1), 74-118.

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.

Ditto, P. H., & Lopez, D. F. 1992. Motivated skepticism: use of differential decision criteria for preferred and nonpreferred conclusions. *63,* 4(568-584).

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

Egan, L. C., Santos, L. R., & Bloom, P. 2007. The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological Science,* 18(11), 978-983.

Elster, J. (1998). Introduction. In J. Elster (Ed.), *Deliberative Democracy*. Cambridge: Cambridge University Press.

Evans, J. S. B. T. 2006. The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review,* 13(3), 378-395.

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

Fiske, S. T., Cuddy, A. J. C., & Glick, P. 2007. Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences,* 11(2), 77-83.

Frantz, C. M., & Janoff-Bulman, R. 2000. Considering both sides: The limits of perspective taking. *Basic and Applied Social Psychology,* 22, 31–42.

Frederick, S. 2005. On the ball: Cognitive reflection and decision-making. *Journal of Economic Perspectives*.

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.

Gouzoules, H., Gouzoules, S., & Miller, K. 1996. Skeptical responding in rhesus monkeys (Macaca mulatta). *International Journal of Primatology,* 17, 549-568.

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.

Hammond, K. R. 1996. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*: Oxford University Press, USA.

Hastie, R., & Kameda, T. 2005. The robust beauty of majority rules in group decisions. *Psychological Review,* 112(2), 494-50814.

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

Koenig, M. A., & Harris, P. L. 2005. The role of social cognition in early trust. *Trends in Cognitive Sciences,* 9(10), 457-459.

Koopowitz, H., & Keenan, L. 1982. The primitive brains of platyhelminthes. *Trends in Neurosciences,* 5(3), 77-79.

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

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.

Krueger, J. L. 2003. Return of the ego-self-referent information as a filter for social prediction: comment on Karniol (2003). *Psychological Review,* 110, 585-590.

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.

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., Kerr, N. L., Davis, J. H., Halff, H. M., & Marciniak, K. A. 1975. Group size, member ability, and social decision schemes on an intellective task. *Journal of Personality and Social Psychology,* 33, 80-88.

List, C. (2004). Democracy in animal groups: a political science perspective. *Trends in Ecology & Evolution, 19*(4), 168-169.

Lerner, J. S., & Tetlock, P. E. 1999. Accounting for the effects of accountability. *Psychological Bulletin,* 125, 255-275.

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.

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.

Mahoney, M. J., & DeMonbreun, B. G. 1977. Psychology of the scientist: An analysis of problem-solving bias. *Cognitive Therapy and Research,* 1(3), 229-238.

Maloney, L. T., Trommershäuser, J., & Landy, M. S. 2007. Questions without words: A comparison between decision making under risk and movement planning under risk. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 297–313). New York: Oxford University Press.

Mercier, H. Submitted. Looking for arguments.

Mercier, H., & Sperber, D. 2009. Intuitive and reflective inferential mechanisms. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds*. New York: Oxford University Press.

Mercier, H., & Sperber, D. Submitted. Why do humans reason? Arguments for an argumentative theory.

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.

Nickerson, R. S. 1998. Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology,* 2, 175-220.

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.

Prescott, T. J. 2007. Forced moves or good tricks in design space? Landmarks in the evolution of neural mechanisms for action selection. *Adaptive Behavior,* 15(1), 9-31.

Prescott, T. J., Redgrave, P., & Gurney, K. N. 1999. Layered control architectures in robots and vertebrates. *Adaptive Behavior,* 7(1), 99-127.

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

Redgrave, P., Prescott, T. J., & Gurney, K. N. 1999. The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience,* 89, 1009-1023.

Rowe, G., & Wright, G. 1999. The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting,* 15(4), 353-375.

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.

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

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., & Nye, P. 1992. The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes,* 51(3), 416-446.

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.

Slovic, P. 1993. Perceived risk, trust, and democracy. *Risk Analysis,* 13(6), 675-682.

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

Spellman, B. A. 1993. Implicit learning of base rates. *Psycoloquy,* 4, 61.

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

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

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

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

Spinoza, B. 1677. *Ethics*.

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

Stanovich, K. E., & West, R. F. 2007. Natural myside bias is independent of cognitive ability. *Thinking and Reasoning,* 13(3), 225-247.

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.

Tindale, R. S., & Sheffey, S. 2002. Shared information, cognitive load, and group memory. *Group Processes & Intergroup Relations,* 5(1), 5.

Toplak, M. E., & Stanovich, K. E. 2003. Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology,* 17, 851-860.

Tversky, A., & Kahneman, D. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review,* 90(4), 293-315.

Wason, P. C. 1966. Reasoning. In B. M. Foss (Ed.), *New horizons in psychology: I* (pp. 106–137). Harmandsworth, England: Penguin.

Whyte, G. 1993. Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes,* 54(3), 430-455.

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., 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.

Yaniv, I., & Kleinberger, E. 2000. Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes,* 83, 260-281.

Zahavi, A., & Zahavi, A. 1997. *The handicap principle: a missing piece of Darwin's puzzle*. Oxford: Oxford University Press.