Collectivized Intellectualism

Benjamin Wald
Julia Jael Smith

Section 1: Introduction

The human capacity to reason is often taken to be responsible for our incredible success as a species. It is natural to assume that the primary evolutionary function of human reasoning is to allow us to gain greater knowledge and from effective intentions, thereby improving our survivability and fitness. Call this the intellectualist view of reasoning. But this natural view seems at odds with the widespread and systematic biases to which our reasoning is subject. Reasoning often seems to do more to entrench whatever we happen to already believe, rather than providing a path to improve or correct our existing beliefs.[1] This has led some to argue that the evolutionary function of reasoning is primarily social, rather than cognitive. Hugo Mercier and Dan Sperber, who propose a view of this kind, call this account of the function of reasoning "interactionism".[2] According to interactionism, reasoning evolved in order to facilitate communication and cooperation, rather than to improve the accuracy of individual's beliefs or the effectiveness of their decision making.[3]

We argue for an evolutionary function of reasoning that combines elements of both the intellectualist and interactionist theories. On our view, which we call collectivized intellectualism, the purpose of reasoning is to increase individuals' knowledge and improve their decision-making, but it evolved to fulfill this purpose when deployed collectively rather than individually. Thus, we agree with the intellectualist that reasoning has a cognitive function, but agree with interactionism that this function is designed so as to be achieved collectively, rather than in isolation.

We start by outlining how collective reasoning could provide cognitive benefits, making it a sensible evolutionary strategy to develop a form of reasoning that exploits the availability of fellow reasoners. Group reasoning is a means of dividing up the cognitive labour of finding arguments and providing error checking

---

[1] See, for example, Nickerson (1998), Kunda (1990) and Ariely (2008) for evidence that reasoning serves to preserve our existing beliefs rather than seeking to correct them.

[2] Mercier and Sperber themselves are the highest profile interactionists (2011, 2017). Andy Norman has proposed a slightly different interactionist picture in response to their work (2016).

[3] For the purposes of this paper, we will be adopting a fairly narrow account of what counts as reasoning. Reasoning is restricted to cases where premises are represented as reasons supporting a conclusion. Thus, cases where a conclusion is reached on the basis of a premise, but without the agent representing that premise as a reason, will not count as reasoning for our purposes. In this, we follow Mercier and Sperber (2017; 2018). One reason for this narrow account of reasoning is that this is plausibly what is unique to humans, and thus the appropriate target for an investigation of the evolution of reason in humans.

on the results. This collective use of reasoning could develop even on a "selfish" interpretation of evolution--there is no need to appeal to controversial processes such as group selection.

Next, we show how our view is superior to both intellectualism and interactionism. The fact that reasoning evolved to be used collectively provides us with resources to explain the puzzling biases that afflict reasoning. While the intellectualist has difficulty accounting for why reasoning would have evolved to feature biases that seem to frustrate its core function, we show that what look like systematic errors in reasoning are in fact explained by its having evolved to be used in a social context rather than in isolation. Reasoning gets the best results when it is used in the context in which it was designed to perform, and falters when it was placed in a context for which it was not designed, like any other evolved trait.

Finally, we consider why one might prefer collectivized intellectualism to interactionism. This is more difficult, since both views predict that reasoning will operate better when deployed in social circumstances, and can offer adaptive explanations for the systematic biases reasoning is subject to. Furthermore, interactionists can agree that reasoning in groups does indeed provide cognitive benefits -- Mercier and Sperber themselves highlight these benefits in the course of their argument for interactionism (2017, Chapter 15). Similarly, the collectivized intellectualist can agree that reasoning does in fact serve a social function, and that this has played an adaptive role. The debate, then, is over which of these is the function that explains the evolution of reasoning, and which is a secondary function that emerged later. This is a difficult question to answer, since it relies on facts about the distant evolutionary past of our species. We offer an argument that the function ascribed by collectivized intellectualism is more likely to have been the first adaptive function of reasoning by arguing that situations in which reason would play the role ascribed to it by the interactionist would plausibly be quite rare in the social context in which early humans evolved, while reasoning would have many opportunities to play the role we ascribe to it. This provides at least a *prima facie* reason to think that our proposed function is the primary function, with the interactionist function arising later as conditions changed to render occasions for reason to achieve this function more common.

Section 2: Collectivized Intellectualism

Our thesis concerns the evolutionary function of reasoning. The evolutionary function of a trait is the effect of that trait which is responsible for its having been selected for by natural selection. The heart, for instance, produces many effects. It makes a loud noise, it circulates blood throughout the body, it consumes energy and generates heat, and so on. Presumably, however, it is the ability of the heart to circulate blood that is responsible for the greater reproductive success of creatures with hearts, and hence this is the evolutionary function of the heart.

It is worth noting a few features of evolutionary functions. First of all, they are historical judgments. What matters is the purpose the trait served when it first became widespread—it may well have been co-

opted to serve new purposes later in the creature's evolutionary history, but this will not change the evolutionary function.[4] Also, the evolutionary function need not be the purpose the organism itself has in using the trait. The evolutionary function of the human sex drive is presumably to create offspring, but that is not the purpose of most human sexual activity. So in asking about the evolutionary function of reasoning, we are asking what it is about the human capacity to reason that improved the reproductive fitness of ancestral humans, or perhaps the fitness of other hominids who first developed reasoning and from whom we are descended (Cf. Cosmides & Tooby, 1997).

Because judgments about evolutionary function are historical judgments, to understand the evolutionary function of a trait we must also understand the environment in which it evolved to function. Both fins and legs have the evolutionary function of locomotion, but each is obviously optimized to fulfill its function in different contexts, and will perform its function quite poorly in the wrong context. As Mercier and Sperber point out, what looks like a flaw in an evolved trait may just be be a result of exercising that trait in an abnormal context (2017, p. 10). Our view, then, is a combination of a claim about the evolutionary function of reasoning with a claim about the environment within which reasoning evolved to achieve this function. We agree with the intellectualist that the function of reasoning is to improve individual cognition, but we argue that the context in which reasoning fulfills this function is collective rather than individual, as is argued by the interactionist.

We clearly do sometimes deploy reasoning in isolation, constructing explicit arguments or comparing the reasons for and against proposed beliefs or courses of action -- for example, when we plan the most efficient route to an unfamiliar destination. However, we also frequently engage in reasoning as a group. Group reasoning has an adversarial dimension, with each member of the group using reasoning to persuade others to adopt his or her own preferred belief or course of action. But this adversarial aspect of group reasoning is compatible with collaborative forms of collective reasoning where individuals share ideas and arguments and try to reach a collective decision that may not match the starting point of any of the participants (Cf. Mercier et. al., 2017).

Even compared to our close evolutionary cousins, the great apes, humans display an extremely high degree of cooperation (Cf. Burkart et. al. 2014). Human hyper-cooperation is a major component of our success as a species, and there is increasing evidence that our cognition across a variety of domains is designed to improve our cooperation. Chris Frith has argued that metacognition, and perhaps even consciousness itself, developed as a tool for allowing us to form more accurate beliefs about the world and about ourselves by discussing with others our beliefs and experiences (2011, 2012). For example, one study

---

[4] There is some debate within the literature on evolutionary function over whether function should be identified with the adaptive advantage that originally gave rise to a trait, or with the adaptive advantage that sustained the trait over more recent evolutionary history (See, for instance, Gould and Vrba, 1982; and Godfrey-Smith, 1994). For example, feathers are believed to have originally served the function of thermoregulation in dinosaurs. Should we then say that the function of the feathers of modern birds is also thermoregulation, or instead is it flight, the function that has been responsible for birds continuing to possess feathers over their more recent evolutionary history? But in either case, a judgment about the evolutionary function of a trait is a historical judgment about the adaptive advantage the trait provided to the ancestors of the modern organism, and which timescale is relevant will not make a difference for our purposes.

shows that two people who collaborated on a visual task, trying to identify when a visual cue appeared on a screen, did better than the better of the two individually, even without being given feedback about when their reports where accurate. Furthermore, the most successful pairs worked by developing a shared vocabulary of metacognitive reports, such as "very sure, sure, uncertain" to report their confidence in their visual judgment (Bahrami et al., 2010, 2012).

Our proposal is that reasoning also developed as a tool to allow us the cognitive benefits afforded by cooperation. Whereas metacognition allows us to cooperate in forming accurate judgments in areas in which we lack access to the cognitive processes underlying our judgments, such as visual judgments, reasoning is an additional tool we can deploy in cases where we have cognitive access to relevant facts that bear on the judgment. Our claim, then, is that reasoning developed to be used in a collective context, to allow us to take advantage of the availability of other deliberators. Importantly, this implies that the design of our reasoning abilities may not be such as to optimize reasoning in isolation. The best way for an individual to contribute to a collective deliberation may not be, and we will argue isn't, the same as the best way for that individual to reason on their own.

The first step in establishing our claim is to demonstrate that reasoning in groups does indeed provide better outcomes than reasoning alone. There is an increasing amount of research that shows that reasoning in groups has a cognitive payoff for individuals, resulting in more accurate beliefs than individuals could have attained through solitary reasoning. For example, individual performance on simple reasoning tasks like the Wason card selection task improves dramatically when people are given the opportunity to solve the problem in groups.[5] When reasoning individually, people perform very poorly on this task, coming up with the correct conclusion about 20 percent of the time (Cosmides, 1989; Wason, 1966). But when reasoning in a group, people arrive at the right answer much more often – 75 percent of participants came to the correct conclusion in a study by Moshman and Geil (1998). Group deliberation also improves individual performance in tasks involving numerical estimations (Minson, Liberman, & Ross, 2011); complex problem-solving tasks (Laughlin, Hatch, Silver, & Boh, 2006); and, in children, in Piaget's conservation tasks (Doise, Mugny, & Perret-Clermont, 1975; Silverman & Geiringer, 1973). Not only is collective decision making more reliable than individual decision making, there is research suggesting that groups can reach a decision on certain tasks as quickly as individuals, even when required to reach a unanimous decision, and still do better than an individual (Blinder and Morgan 2000).

What explains this improved success? There are likely to be a range of mechanisms, but we want to suggest two. One way in which reasoning in a group context benefits the individual is by allowing him to share cognitive labour with his fellow deliberators. When faced with a difficult decision that involves identifying considerations for and against different courses of action, individuals are limited by time and

---

[5] In the Wason card selection task, participants are shown four cards placed on a table. Each card has a number or letter printed on the visible side. Participants are told that each card has a letter written on one side, and a number on the other. Then, they are asked which cards need to be turned over to test whether the rule '*if there is a vowel on one side of a card, then there is a even number on the other side*' is true of the four cards. Although the participants have enough information to obtain the correct answer through logical reasoning, repeated experiments have shown that people perform abysmally on this task, getting the right answer 25% of the time or less (Wason, 1966).

working memory. Reasoning in groups allows us to split up the task of identifying arguments for different courses of action, with each member of a group performing a "depth first" search, focusing on a narrow range of arguments that they explore in detail. When group members share these arguments with others, they can survey the arguments for different positions without wastefully duplicating the cognitive labour of their fellow deliberators. Mercier and Sperber identify and elaborate on this benefit of group deliberation, noting that reasoning in groups can give us easier access to considerations for and against different conclusions than we would have identified alone (2017, p. 221).[6]

There is a wealth of experimental evidence of the improved performance of human reasoning under these conditions, supporting the idea that there is such a division of labour and that it leads to improvements in individual's knowledge and decision-making.   Another benefit of group deliberation is error checking. Reasoning is a sequential process. Each stage of reasoning provides its output as input to the next stage -- indeed, this has been said to be one of the defining features of reasoning (Evans & Stanovich, 2013, p. 235). This gives reasoning great power. Chains of reasoning can lead us from modest starting points to surprising and novel end points. However, there is a correlative risk to this iteration. Small errors introduced early on will be carried along and may be magnified at each stage, potentially leading to much larger and more dangerous errors than would be possible in a single stage of reasoning. In order to reason about complex topics, we need to create long chains of reasoning. When we reason individually, it is easy to miss small errors early in this reasoning, and so to end up with wildly wrong conclusions. On the other hand, when we share our reasoning with others, each other person provides a form of error checking, greatly reducing the chances of errors propagating through our reasoning. It is simply a fact about cognitive systems that it is difficult for them to check their own functioning, while having external systems perform such a check improves our reliability, since the chance of both systems having an identical malfunction is low. This is an effective solution, as shown by the fact that we use a similar error checking system in computers, having independent machines perform the same calculation (e.g. Mohat, 1998). Reasoning in groups provides the benefit of having others to catch our errors.

These natural advantages to collective deliberation explain why, despite having an individualistic function, reasoning could nonetheless have evolved to operate best in a collective context. We want to stress that our view does not need to appeal to controversial evolutionary forces such as group selection.[7] It is clearly possible for an activity that is best performed collectively to serve an evolutionary function

---

[6] On some definitions of reasoning, generating arguments is not itself a part of the reasoning process per se, but a precondition for reasoning. We need some distinction between pre-conditions for reasoning and reasoning proper-- the fact that we reason better on a full stomach does not make eating a sandwich part of reasoning, even if it improves the accuracy of the result. But for the purposes of investigating the evolutionary origin of reasoning, it makes sense to us a broader definition of what is part of reasoning. The adaptive benefit of reasoning will rely not only on how well we draw logical entailments from our premises, but also how and when we use this ability, which requires us to consider what premises we use and what tasks we use reasoning for. Thus, we feel comfortable including these as parts of reasoning for our purposes.

[7] Group selection is the claim that evolutionary pressure can, in some contexts, operate at the level of the groups rather than the individual, so that traits harmful for the individual but beneficial for the group might nonetheless be favoured by evolution. While there has been some recent defenses of the claim that evolution can operate at the level of groups, our own claim need make no such claim.

aimed at individual benefit. Consider pack hunting behavior, such as that shown by wolves. Wolves will engage in group hunting tactics that rely for their success on the contribution of other wolves. This allows them to catch larger prey than they could individually. The evolutionary function of this collective behavior is individualistic--participating in the group hunt increases fitness insofar as each wolf that takes part in the collective hunt procures food for itself and its offspring. There is no group selection needed to explain this result, just the fact that individual wolves who participate in pack hunting will survive in greater numbers than their counterparts who hunt alone.

Dessales (2011) argues that understanding argumentation as a means of improving overall beliefs will need to rely on some form of group selection. If communication improves the beliefs of the speaker, he argues, then it fails to benefit the hearer, and if it improves the beliefs of the hearer than it fails to benefit the speaker. In either case, someone is expending effort, in transmitting or correcting beliefs, without gaining any benefit. But this omits the idea that a collective endeavor could provide results that are superior to the starting beliefs of any of the participants. Reasoning as a group need not be a mere transmission of accurate information from those who possess it to those who lack it--it can be a way of generating new information for the group as a whole that is more accurate than the contribution of any of the individual group members.

Thus, just as collective hunting provides more food per wolf than individual hunting would, collective reasoning provides each reasoner with a reward of more accurate beliefs for less cognitive effort than they could have had by reasoning alone. Of course, this kind of collective endeavor might need to evolve mechanisms to discourage free-riders from benefitting without contributing. But it is plausible that collective reasoning in humans possesses just such mechanisms to reward those who contribute most and punish those who shirk. There are important reputational gains to being seen as wise and helpful in group activities, including reasoning (Cf. Cosmides, 1989), and costs to being seen as uncooperative or unintelligent (Cf. Boyd et. al. 2010). Thus, a form of social selection could ensure that cooperation was evolutionarily favoured.

Of course, there is a cost to collective hunting that individual hunting avoids: the wolves must share the food. There would therefore be a benefit to each individual wolf if it could skip involving other wolves and just take down the large prey itself. So why hasn't evolution given each individual wolf the speed and strength to hunt alone? The details we leave to biology, but the general form of the answer is apparent. This degree of speed and strength for wolves is either not evolutionarily available, given the genetic and physical constraints, or would entail unacceptable tradeoffs in other areas. Similarly, being capable of reasoning individually would be helpful, because sometimes we might not have any appropriate interlocutors to engage in group deliberation, and because reasoning with others takes time and may involve reputational costs. But presumably such prowess at individual reasoning is either not practical, or else would require unacceptable tradeoffs in other areas.[8]

---

[8] We might wonder how a trait like collective reasoning could have originally evolved. A mutation must originate in a single individual, but the propensity to reason collectively will be useless until it is sufficiently widespread for there to

Collective deliberation, then, is like collective hunting. It is an activity that humans do better as a group, but its evolutionary function is to improve an individual's knowledge and decision-making, with the social functions coming along as a welcome side-effect. Furthermore, in the following sections we will argue that reasoning has some of the weaknesses it has precisely in order to allow better group deliberation, equivalent to a pack hunting animal that has evolved to be better at group hunting in ways that render it less suited to hunting alone.

The best way to demonstrate the plausibility of this view is to compare it to the competing accounts of the evolutionary function of reasoning. In the following two sections, we'll consider two competitors to our view: the dominant intellectualism, and Mercier and Sperber's interactionism. We will show that our own view has significant explanatory advantages relative to each of these alternative accounts.

Section 3: Intellectualism

The assumption in most of the philosophical and psychological research on reasoning has been that reasoning evolved not just to improve individual beliefs and decision making, but that it evolved to perform this task individually. This is a natural starting assumption. A capacity that can be deployed in isolation will be more robust since it will not depend on either the presence or cooperation of others to yield its benefits. We will follow the lead of Mercier and Sperber, who dub this view of the evolutionary function of human reason the *intellectualist* view (2017, p. 4).

Why would we consider rejecting standard intellectualism in favour of a collectivized version of the view? Over the past fifty years, experimental work in psychology has been accumulating evidence that demonstrates that reasoning is subject to numerous biases and failures. And these biases do not appear to be merely due to a lack of reasoning ability or training; instead, they seem to be systematic biases that incline us towards supporting our own views and criticizing views we disagree with. This confirmation bias is an entrenched feature of the human reasoning ability, but it is not the kind of feature that a system designed for solitary thinkers to improve their judgments would possess.

In seeking an explanation for these failures, one possibility is that such errors in reasoning are a result of chance: they are simply mistakes that result from our limited capacities as human animals. Naturally, there are some cases in which human reasoning will fail to secure knowledge for the individual, but the intellectualist might argue that this is what one would expect, given the complexity of many reasoning problems and our fallibility. As such, the intellectualist might say, the data doesn't count as a sufficient reason to reject intellectualism.

---

be partners to reason with. This is a general issue with the evolution of any cooperative trait, and the widespread existence of cooperation both within species, as with pack hunting or eusociality, and between species, in mutualistic relationships such as those between some species of ants and aphids, shows that this barrier is not insurmountable. Presumably a number of individually beneficial mutations set up the conditions for some kind of proto collective deliberation, which then sets the stage for mutations that are adaptive specifically because of their contribution to improving collective deliberation.

This response, however, will not work. Not only is human reason not always effective at attaining the intellectualist goals, but it has features that seem directly detrimental to those goals. Take the case of confirmation bias. People systematically use reasoning not to challenge or alter their existing beliefs, but to confirm them. People use reasoning to find reasons for what they already believe, or to find reasons for rejecting what they already disbelieve (See Klayman (1995) and Nickerson (1998) for an overview of the collection of phenomena that are typically characterized as forms of confirmation bias). Using reasoning in this way fails to reliably increase the accuracy of beliefs or the appropriateness of intentions. Indeed, this is the opposite of the tendency we would expect in an adaptation meant to improve the accuracy of beliefs. We would do better to seek out reasons to doubt or revise our beliefs, since this would be the most effective at uncovering errors and producing improvement in our cognition. Indeed, this is what we see in other faculties designed to furnish us with accurate beliefs. Vision, for example, has a bias towards attending to the unexpected or surprising, which makes sense as an adaptation to quickly detect and correct errors in our perceptual beliefs (Summerfield & Egner 2011).

The intellectualist might hope that confirmation bias is the result of flaws in individual reasoners, rather than in reason itself, and that those who are better at reasoning would be less affected by confirmation bias. Unfortunately, there is evidence that suggests that those who are better trained in how to reason and who tend to use reasoning more often are more, not less, likely to display this bia. Dan Kahan, for example, points to evidence that conservatives who have better reasoning skills and more information on climate change are more likely to deny anthropogenic climate change than conservatives who lack such skills (2015). The most plausible explanation is that the better someone is at reasoning the more easily they are able to come up with rationalizations for their prior opinions, and the less likely they will be to find arguments on the other side persuasive.[9]

This kind of feature of reasoning is difficult to dismiss as mere 'user error'. Ordinarily, we would expect a lack of reasoning capacity to produce errors more or less at random. But confirmation bias has a clear direction—the error is always in favor of one's own existing beliefs. This suggests that the bias is a feature of reasoning, not a mere side effect of imperfect reasoning ability. Since confirmation bias makes no sense as an adaptation to improve isolated reasoning, the intellectualist will need to try to explain this feature as a side-effect of some other adaptation, and, as Mercier and Sperber point out (2017, p. 281), there is no obvious explanation of this kind available for the intellectualist.[10]

_____

[9]See Isenberg (1986) for an overview of further evidence for this explanation of group polarization.

[10] We claim that the existence of confirmation bias thwarts the intellectualist's story about the function of reasoning. Another way for the intellectualist to resist this conclusion is to argue that confirmation bias isn't really a bias, but is in fact a reasonably effective means of attaining truth in most ordinary circumstances. (An example of this kind of strategy: Klayman and Ha (1987) argued that the tendency to employ positive search strategies by looking for confirming instances of a hypothesis rather than disconfirming ones is an effective way to test a hypothesis in many ordinary circumstances.) However, as we have already mentioned, 'confirmation bias' is an umbrella term for many disparate phenomena that show that, in general, individuals tend to place more weight in their own beliefs than would be optimal from the perspective of attaining true beliefs. It is unlikely that our tendency to employ positive search strategies can account for all the ways in which we are prone to give undue weight to our own opinions and abilities. Therefore, if the intellectualist were to pursue this strategy, she would have to show not only that positive search strategies tend to produce accurate beliefs, but that other phenomena such as the primacy effect, belief persistence,

Our view, on the other hand, can show how what seems to be a bug is actually a feature. As we discussed above, one of the advantages of collective reasoning is the ability to divide up the cognitive labour of discovering and assessing arguments. Individuals' being prone to confirmation bias would allow them to make the most of the division of cognitive labour in the context of group deliberation. Being inclined to give slightly more weight than is due to our own views encourages us each to build the best case we can for one particular position (our own). Then, as long as the group features a variety of starting positions, we will all get the benefit of seeing the best case for a variety of different views, without each having to work out the arguments for all of the views by ourselves (Cf. Mercier and Sperber 2017; p. 218-221).[11]

In a similar vein, Neil Levy (2017) argues that our individualistic biases confirm the hypothesis that humans are adapted for collective deliberation because of how naturally these biases secure cognitive advantages when employed in group reasoning. Levy points out that we are 'epistemic individualists', disposed to be overconfident in our own ability to reason to accurate conclusions. But our individualistic tendencies pay off when they are used within the context of group deliberation. One benefit of overconfidence in our own powers of solitary deliberation is that tendencies towards individualism among members of a deliberating group helps to ensure that individuals' reasoning is independent of other group members'. Another benefit is that individualists will be more likely than those who are less individualistic to pursue in depth novel explanations or hypotheses. Indeed, pioneering work requires a good deal of self-confidence, continuing with a chain of reasoning even when other group members take it to be flawed (Cf. Kitcher, 1995). Our individualistic tendencies impair our ability to reason well on our own, but they constitute an effective means of coming to form accurate beliefs in the context of collective deliberation. We agree with Levy that the fact that individualistic bias contributes to the remarkable success of group deliberation suggests that reasoning was adapted to secure greater knowledge when used in group contexts.[12]

If the goal of reasoning were to give individuals more accurate beliefs when reasoning in isolation, then confirmation bias, and other of our individualistic tendencies, would be evidence that reasoning is failing miserably at its intended function. But even more worryingly, the failures of reason do not seem to be due to the complexity of the issue or to a lack of skill of reasoners; it seems to be due to a fundamental feature of human reasoning. The fact that reasoning has features that *systematically* impairs individuals' efforts to improve their knowledge through reasoning is is tension with the traditional intellectualist view of reasoning. However, confirmation bias in individuals is adaptive given a collectivized intellectualist account.

---

and preferential treatment of evidence that support one's existing beliefs are *also* beneficial from the point of view of pursuing truth. We think that pursuing this strategy is unlikely to be fruitful.

[11]Tindale and Sheffey (2002) and Meyer and Scholl (2009) provides experimental evidence that groups benefit from a division of cognitive labour. Interestingly, they also suggest some degree of informational overlap is also helpful--it is useful for group members to have overlapping but not identical sets of information in forming group decisions.

[12] Our argument is compatible with the possibility that reasoning first evolved to be of benefit to individual reasoners, but then over time traits that benefited collective deliberation at the expense of individual deliberation were selected for. As discussed in footnote 5, depending on which account of evolutionary function we prefer this may or may not affect what we identify as the evolutionary function. But in either case, the bulk of our argument remains unaffected--evolutionary pressures adapted human reasoning to function better collectively at the expense of individual deliberation.

Our individualistic tendencies (including confirmation bias), then, support the conclusion that reason evolved in order to make us effective collective deliberators.

Section 4: Interactionism

While the intellectualist account is the traditional view of the function of reasoning, it has recently been challenged. In *The Enigma of Reason*, Mercier and Sperber argue for an *interactionist* account of human reason, according to which the function of reasoning is to "overcome problems of trust in coordination and in communication that reduced the ability of early humans to properly benefit from their cognitive and social resources" (Sperber and Mercier, 2018, p. 1). The interactionist sees reason chiefly as a social tool that evolved to facilitate cooperation and the sharing of important information between people. The interactionist views argumentation as valuable because it is a means for us to convince others to adopt our beliefs and join in on our projects when those others cannot rely on us to be trustworthy or benevolent towards them.

To see the value of reasoning ability for the interactionist, let us imagine a scenario that might have faced our distant ancestors. The anachronistically named Jim knows of a large windfall of food. Harvesting all of this food would require the labor of two people, so Jim wants to convince Sally to help him. He could just tell her what he knows, but she may not be willing to believe him just on his say-so. After all, if Jim is mistaken or lying, then Sally stands to waste time and effort, and perhaps even be put in danger. Sally doesn't know Jim well enough to decide on independent grounds if he is trustworthy. But on the other hand, if she just disbelieves Jim then she misses out on an opportunity for mutually beneficial cooperation. Thus, it is in both Jim and Sally's interest for Jim to have some way of demonstrating the truth of his claim that does not depend on an assessment of his trustworthiness. Argumentation, Mercier and Sperber claim, provides just such a mechanism.

Jim can back up his credibility by providing arguments in support of his claim. Importantly, the role of argumentation is to convince others, and thus the arguments Jim provides need not have played any role in generating the belief that Jim is attempting to get Sally to share. Mercier and Sperber argue that our beliefs are usually generated without any reasoning, coming instead from sub-personal modules of intuitive inference (2017, Chapter 9). The inner workings of these modules cannot easily be shared with others--they do not give us conscious access to the grounds of the belief that is their output. This is why we need a separate ability to generate arguments for the belief in order to convince others to share it, and thus get the benefits of cooperation.

Since the goal of reasoning is to get others to share beliefs I already possess, it makes sense that reasoning is biased towards coming up with arguments that support my existing beliefs, regardless of whether those beliefs are accurate. Thus, confirmation bias is a sensible adaptation, given this conception of the goal of reasoning. And it also makes sense that, given interactionism, we would evolve to be epistemic individualists. Overconfidence in my own ability to reason provides me with motivation to invest energy in producing reasons for my beliefs, and motivation to share those reasons with others, serving the

evolutionary function of producing arguments to convince others. Thus, Mercier and Sperber argue that the experimental data showing the purported 'failures' of reasoning is as we would expect, given an interactionist view of human reason: reasoning seems much better suited to achieve the social function of facilitating cooperation than it is suited to achieve the cognitive function of improving our knowledge. The interactionist view, then, provides a more attractive view of the evolutionary function of reason than the traditional intellectualist view.

We might worry, however, that for the interactionist there will be no explanation for why we try to offer good arguments. If our goal is just to convince others, shouldn't we be indifferent between sound arguments and sophistry, as long as the argument that we employ achieves our goal? And if we are indifferent between sound arguments and sophistry, then the interactionist encounters a further problem in explaining the spectacular outcomes in cases of group reasoning discussed in the previous section. If a good argument is merely persuasive rather than sound, why should we expect to see the connection that we clearly *do* see, at the group level, between reasoning and knowledge?

Mercier and Sperber have an answer to this objection. On the interactionist view, the explanation for why reasoning with others improves our performance with respect to the intellectualist goals of knowledge acquisition and decision-making is that others have evolved mechanisms of 'epistemic vigilance' geared towards spotting untrustworthy individuals and false testimony (2017, p. 191-4; Sperber et. al., 2010). From Jim's point of view, it is irrelevant whether the arguments he offers are sound. All he cares about is convincing Sally. Sally, on the other hand, cares very much that the arguments be sound. She wants to avoid being duped or mislead. So it is fitness-enhancing for receivers of arguments to accept only good arguments, so as to avoid both being deceived (if they accept bad arguments), and missing out on valuable opportunities for cooperation (if they reject good arguments). In other words, Mercier and Sperber predict that people will be picky about what kinds of arguments they accept--and the experimental data seems to bear this out. People are better at detecting flaws in reasoning for positions they do not themselves support (Lord & Taylor, 2009). But given that hearers will only accept good arguments, this means that speakers need to learn to provide good arguments if they want to be successful in convincing others. Thus, on the interactionist account, evolution would have selected for those who could offer strong arguments, since this is the only kind of argument likely to be accepted.[13] The fact that we have evolved mechanisms of epistemic vigilance is what allows the interactionist to explain the connection between knowledge and reasoning in group deliberative contexts.

It's important to recognize that this story is not primarily one of competition. Andy Norman (2016), for example, objects to Mercier and Sperber's account on the grounds that if reasoning served only the interest of the speaker, and not the listener, then listeners would quickly learn to disregard attempts to convince them. But Mercier and Sperber do not assume that the benefit of a successful persuasion flow

---

[13] This is not to deny that there could be other systems of convincing others also at work. For example, there may be systems designed to detect trustworthiness based not on arguments but on things like facial cues and body language (the kind of signal often faked by con men), and this could also work to facilitate cooperation. Mercier and Sperber just need to claim that argumentation is one system that facilitates cooperation and communication.

solely to the persuader. In the main case, the person doing the persuading is not a con man (or, as Norman puts it, a used-car salesman), but a hopeful collaborator seeking mutually beneficial cooperation. This is why it is worthwhile for others to listen--reasoning is a tool for finding the potential cooperators and avoiding the con-men. Thus, in the paradigm case of reasoning, both speaker and listener stand to benefit from a successful use of reasoning to persuade.

The interactionist, then, has a straightforward explanation for data revealing the purported flaws of reasoning: they aren't flaws. When the primary goal of reason is to give us tools to communicate and cooperate with others, and to enhance our reputation and credibility through persuasion, cases in which isolated human reasoning fails to improve an individual's knowledge or decision-making come as no surprise. In these cases, we have no reason to expect reasoning to function properly, as it is being used outside of its intended design parameters, for a purpose that it was never meant to achieve.

However, as we argued in the previous section, collectivized intellectualism can also explain this same data. According to collectivized intellectualism, reason evolved to bring cognitive benefits to the individual, but that in order to overcome limitations on an individual's reasoning powers, human reason evolved in such a way so as to exploit the advantages associated with group deliberation. These same adaptations that improve the outcome of collective deliberation make us worse at individual deliberation. We have argued that this evolutionary trade-off between being better at individual deliberation and being better at contributing to a collective deliberation provides an explanation of the data that shows the failures of reasoning in isolation.

So far, we have showed that both collectivized intellectualist and interactionist accounts can explain systematic errors in individual reasoning *and* improved performance in group reasoning: both mesh with what we know about reasoning from the psychological literature. According to the interactionist, though reasoning evolved to help individuals to position themselves socially by offering justifications and persuasive arguments, in order to achieve this end, it was necessary that successful reasoning *also* further the intellectualist goals of greater knowledge and improved decision making for the individual members of a reasoning community. If this is correct, the claim that reason's *primary* function is social is a claim about the evolutionary history of reasoning in our species: Interactionism holds that the trait that was first selected for because of its fitness enhancing properties was the ability to persuade others through reasons-giving behaviour, with cognitive benefits coming along as a helpful side-effect.[14]

The collectivized intellectualist account, on the other hand, posits that the truth-seeking function, not the social function, of reason is primary. This amounts to the claim that reasoning's ability to improve individual knowledge and decision-making when exercised in a group context was the fitness-enhancing feature that was first selected for in the evolutionary history of human reasoning. Collectivized intellectualism results in a different explanation for the success of group reasoning: that we arrive at greater

---

[14]Mercier and Sperber note that, when forming hypotheses about primary function, "the challenge … is to establish what beneficial effects would explain why reason evolved at all" (2017, p. 200). In other words, M & S take establishing the primary function of a trait to involve determining what features of that trait allowed it to first propagate throughout a population.

knowledge when we reason in groups because human reason evolved for this very purpose. On the collectivized intellectualist account of reason, it might make sense to think of the social benefits of reasoning -- namely, improved cooperation and a tool for facilitating trust between individuals -- as a kind of secondary function.[15]

Due to the historical nature of judgments about evolutionary function, we cannot easily answer the question about which is the primary and which is the secondary function in evolutionary terms from investigation of modern reasoners. Instead, we need to look to the distant past, and to the environment and circumstances in which our early ancestors developed the capacity to reason. Modern humans are genetically virtually identical to our immediate ancestors who lived 150,000 years ago. So the relevant context for the evolution of reasoning is not modern civilization, but the scattered hunter-gatherer groups of our ancestors.

To see which proposed function of reasoning is a better candidate for its primary evolutionary function, consider the relative frequency with which each proposed function would have come up in the environment in which reasoning first evolved. How often would our early ancestors have had the opportunity to have collectively obtained new knowledge and made good decisions as a group? How often would they have had the opportunity to persuade others to adopt their own beliefs? Given what we know about the daily lives of our ancestors, opportunities for collective knowledge-seeking and decision-making through reasoning would have been more frequent, and would have had greater payoffs, than opportunities for overcoming the distrust of others through reasoning. If this is so, collectivized intellectualism is more likely to accurately describe the primary function of reasoning than interactionism.

Of course, it is difficult to know exactly how our distant ancestors lived. But anthropologists have argued that their lifestyle probably most closely resembled those of modern hunter-gatherer tribes, such as those observed in Africa (see, e.g., Marlowe 2005; White et. al., 2003). These groups are quite small, with only a few hundred individuals at most, and interactions with total strangers would therefore be comparatively rare. Thus, exchanges of information and opportunities for collaboration would tend to be within the group. But Mercier and Sperber's evolutionary story relies on the idea that reasoning is used in contexts in which we want to share information, but the recipient lacks the ability to assess the trustworthiness of the speaker. This comes up frequently in modern society--we often interact with total strangers, whose competence and goodwill we have no prior knowledge of. But it is implausible that this kind of meeting would occur frequently in the hunter-gatherer context. Contact with strangers would be

---

[15] Several people have suggested to us the possibility of a 'dual function' view, according to which the evolutionary function of reasoning is both social and cognitive. Presumably, this kind of a hybrid view would hold that, in the context that reasoning first evolved, the social and the cognitive advantages that accrued from the ability to reason were equally operative in the spread of reasoning throughout the population. So far, we've argued that, given what we know about the advantages of group reasoning and the pitfalls of isolated reasoning, it is still highly plausible that human reasoning has a cognitive evolutionary function (*contra* the interactionists, who claim that this data shows that the primary function of reasoning is *not* cognitive). Our argument does not preclude the possibility of a hybrid view, and we are not opposed to such a view. However, we also think that there is some reason to prefer the collectivized intellectualist view of function over the interactionist view, as we explain in the remainder of the paper.

rare--one would have some familiarity with all the members of one's own group, since it would be a relatively small number of people. Anyone you are communicating or cooperating with is probably someone you have cooperated with in the past and will continue to collaborate with in the future. In this kind of ongoing relationship, the need to determine the truth of testimony in absence of knowledge of the speaker's reputation is greatly reduced. After all, if someone in your community lies to you, they can be punished by refusal to collaborate in the future. With a small pool of potential collaborators, lying to secure cooperation is unlikely to be viable strategy, since one will rapidly be excluded from future collaborations. This is not to say that the kind of scenario envisaged by Mercier and Sperber would never occur--perhaps it would occur in encounters between bands or tribes of hunter-gatherers. But these kinds of interactions are unlikely to have been frequent or important enough to support the development of a unique tool like reasoning just to resolve them.

Even in interactions with someone with which you are familiar, a means of assessing their truthfulness on a particular occasion could still be useful. The system of trust backed up by punishment has the disadvantage of only operating after the fact, and that punishments often incur costs to both the punished and the punisher. The risk of punishment can thus fail to ensure truthfulness in some very high stakes and some very low stakes situations. High-stakes situations because the benefit of betraying the trust may be larger than the costs of forsaking future interactions, and low stakes cases because the cost to the victim of punishing the untrustworthiness may outweigh the benefits of avoiding future untrustworthiness. An example of a case of the first kind would be a criminal who betrays his confederates by running out with the profits from their once-in-a-lifetime heist; the benefit of the one-time payoff is bigger than the cost of never working with those confederates again. An example of the second kind is if one friend never picks up the cheque; the cost of losing their friendship may outweigh the cost of always paying for coffee, so they avoid punishment.

These kind of cases show that even in a small hunter-gatherer group there may be some benefit to the interactionist function of reasoning that trust and punishment couldn't provide. However, neither kind of case seems a plausible candidate for being the original evolutionary function of reasoning. The high stakes cases are simply too rare to be a major component of the fitness of ancestral humans. It's hard to see what kind of situation would come up in which a one-time benefit would be large enough to compensate for being ostracized from one's community. The small stakes cases, on the other hand, are likely to be common but by definition they involve small stakes. Thus, if they were the primary driver of the evolution of reasoning, then reasoning would be an adaptation that just served to fill in a few of the gaps in the system of trust and punishment that did most of the work to facilitate cooperation. While possible, reasoning is such a complex adaptation it seems unlikely to have been driven by such a relatively minor fitness advantage. Instead, it is more likely that this would be a secondary benefit of reasoning that had evolved to provide some other more substantial fitness benefit, such as collective deliberation.

So the interactionist function of facilitating cooperation would have contributed only slightly to the wellbeing of people living a hunter-gather lifestyle. Collective reasoning, on the other hand, would be

likely to be highly beneficial to hunter-gatherers. Much of the activity of such groups is collective already, and major decisions tend to be made collectively.[16] Group reasoning could have helped decide where to look for food, when to move on from an area, and so on. Collective reasoning would be helpful under a range of circumstances that would occur very frequently, whereas arguing as means of convincing others to trust us would be helpful only under specific, infrequent circumstances. This suggests that collective reasoning is more plausibly the primary function of reasoning, with the ability to transfer information serving as a helpful secondary function that has increased in importance over time as interaction with strangers has become more common and more important to our lives.

References

Ariely, D. (2008). *Predictably Irrational.* New York: HarperCollins.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science, 329*(5995), 1081-1085.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012, 02). Together, slowly but surely: The role of social interaction and feedback on the build-up of benefit in collective decision-making. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 3-8.

Blinder, A. S., & Morgan, J. (2000). Are two heads better than one?: An experimental analysis of group vs. individual decision making. NBER Working Paper No. w7909. Available at SSRN: https://ssrn.com/abstract=242143

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science, 328*(5978), 617-620.

Burkart, J.M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., Huber, J., Isler, K., Kosonen, Z.K., Martins, E. and Meulman, E.J.(2014, 08). The evolutionary origin of human hyper-cooperation. *Nature Communications, 5*, 4747. doi:10.1038/ncomms5747

Cashdan, E. A. (1980). Egalitarianism among hunters and gatherers. *American Anthropologist*, *82*(1), 116-120.

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: A primer.* Center for Evolutionary Psychology, Santa Barbara.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason?

---

[16] Typically, the societies of hunter-gatherers are 'egalitarian' rather than hierarchical: there are no relationships of dominance within same-sex, same-age groups (i.e. there are no dominant adult males and no dominant adult women, though males may occupy a position of greater authority than females, or vice versa) (Fried, 1967). Egalitarian societies are characterized by social practices that discourage displays of arrogance or authority and the hoarding of resources (Cashdan, 1980; Draper, 1978; Henrich et. al., 2004). In a typical egalitarian society, no one individual would have the authority to make major decisions for the group. Thus, typical egalitarian societies represent a context rife with opportunities to realize the cognitive gains of collective reasoning.

Studies with the Wason selection task. *Cognition, 31*(3), 187-276.

Dessalles, J. (2011). Reasoning as a lie detection device. *Behavioral and Brain Sciences*, 34(02), 76-77

Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European journal of social psychology*, *5*(3), 367-383.

Draper, P. (1978). The learning environment for aggression and anti-social behavior among the! Kung.

Evans, J. S., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition. *Perspectives on Psychological Science, 8*(3), 223-241.

Fried, M. H. (1967). *The evolution of political society: An essay in political anthropology* (Vol. 7). Random House.

Frith, C. D. (2011, 06). Consciousness is for sharing. *Cognitive Neuroscience, 2*(2), 117-118. doi:10.1080/17588928.2011.585230

Frith, C. D. (2012, 06). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1599), 2213-2223.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, *28*(3), 344-362.

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology, 8*(01), 4-15.

Henrich, J. P., Boyd, R., Bowles, S., Fehr, E., Camerer, C., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies.* Oxford University Press on Demand.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, *50*(6), 1141.

Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, *36*(S1), 1-43.

Kitcher, P. (1995). *The advancement of science: Science without legend, objectivity without illusions.* Oxford University Press on Demand.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32, 385-418

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review, 94*(2), 211.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498.

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. *Journal of Personality and social Psychology*, *90*(4), 644.

Levy, N. (2017, 06). Due deference to denialism: Explaining ordinary people's rejection of established scientific findings. *Synthese.*

Lord, C. G., & Taylor, C. A. (2009). Biased assimilation: Effects of assumptions and expectations on the interpretation of new evidence. *Social and Personality Psychology Compass, 3*(5), 827-841.

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 14*(2), 54-67.

Mercier, H., Boudry, M., Paglieri, F., & Trouche, E. (2017). Natural-born arguers: teaching how to make the best of our reasoning abilities. *Educational Psychologist*, *52*(1), 1-16.

Mercier, H., & Sperber, D. (2011, 03). Argumentation: Its adaptiveness and efficacy. *Behavioral and Brain Sciences, 34*(02), 94-111.

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Meyer, B., & Scholl, W. (2009). Complex problem solving after unstructured discussion: Effects of information distribution and experience. Group Processes & Intergroup Relations, 12(4), 495-515.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325-1338.

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

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

Norman, A. (2016). Why we reason: Intention-alignment and the genesis of human rationality. *Biology & Philosophy, 31*(5), 685-704.

Silverman, I. W., & Geiringer, E. (1973). Dyadic interaction and conservation induction: A test of Piaget's equilibration model. *Child development*, 815-820.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359-393.

Sperber D, Mercier H. (2018). Why a modular approach to reason? *Mind and Language*. 2018;1–9. https://doi.org/10.1111/mila.12208

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *13*(9), 403-409.

Tindale, R. S., & Sheffey, S. (2002). Shared information, cognitive load, and group memory. Group Processes & Intergroup Relations, 5(1), 5-18.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012, 12). Two Key Steps in the Evolution of Human Cooperation. *Current Anthropology, 53*(6), 673-692.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.),*New Horizons in Psychology* (pp. 106-137). Harmandsworth: Penguin.

Wason, P.C.(1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20, 273---281.

White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., & Howell, F. C. (2003). Pleistocene homo sapiens from middle awash, ethiopia. *Nature*, *423*(6941), 742.