Under influence

Frédérique de Vignemont, Hugo Mercier

F. de Vignemont & H. Mercier (draft). Under influence. In *Alvin Goldman and his Critics*, edited by Hilary Kornblith and Brian McLaughlin, Blackwell.

Under influence

Vignemont&Mercier

In many circumstances we tend to assume that other people believe or desire what we ourselves believe or desire. This has been labeled 'egocentric bias.' Egocentric bias seems to be particularly true of young children. For many years, it has indeed been asserted that children under four years of age cannot adopt someone else's perspective, as shown by their failure to pass the false-belief task. Instead, they judge that the others share their true belief. But children are not the only victims of egocentric bias. Even adults, who are supposed to have achieved sophisticated mindreading abilities, can neglect differences between their own perspective and other people's perspective. Experts are particularly likely to suffer from a form of this problem known as the curse of knowledge: for example, business experts can fail to discount their own knowledge when predicting corporate earning forecasts by less-informed people (Camerer et al., 1989). Egocentric bias has been taken as evidence that our own perspective has some kind of priority: (i) developmental priority: we learn first what we believe, and only later can we judge what others believe; (ii) computational priority: it is less cognitively demanding to judge what we believe than what others believe; (iii) epistemic priority: we have a privileged access to our own beliefs that we do not have for other people's beliefs.

One of the assets of the Simulation theory, as defended by Alvin Goldman in many papers and in his recent book *Simulating minds*, is its ability to explain egocentric bias, and more generally the priority of first-person mindreading (self-ascription of mental states) over third-person mindreading (ascription of mental states to other people). In particular, Goldman accounts for egocentric bias in terms of failure to quarantine one's own perspective. When one tries to understand other people, one puts oneself in their shoes. To do so, one pretends to be in the same situation and to have their beliefs and desires. This involves inhibiting one's own beliefs and desires. But one can neglect or fail to do so. This results in egocentric errors.

This is not to say that we systematically fail to understand other people and forget that they can have a different perspective. If it were the case, then it would be highly difficult, if not impossible, to communicate, cooperate or compete with them. In those situations, we need to take the other person's perspective and to inhibit our own. But can the other's perspective

furtively intrude even when no reason seems to require it, or even when it is detrimental for us? We shall see a series of evidence of what has been called *altercentric bias* (Samson et al., 2010; Apperly, 2011): other people's beliefs can unduly influence us even when they are wrong. At first sight, altercentric bias questions 1st person priority. In particular, it may appear as incompatible with simulation-based accounts of 3rd person mindreading. We shall argue, on the contrary, that the simulationist framework enables confusions between self and others that go both ways: taking one's beliefs for the other's beliefs (egocentric bias) and vice-versa, taking the other's beliefs for one's beliefs (altercentric bias). We shall then see how the risk of such confusion may be disadvantageous from an evolutionary perspective, questioning thus the evolutionary plausibility of the simulation theory.

1. When the self takes over

For many years, the debate in the mindreading literature was framed in either/or terms: either 3rd person mindreading results from theorizing or from simulating. Goldman convincingly argued in favor of the latter. By putting ourselves in the others' shoes and running off-line our own cognitive resources, we can simulate or re-create other people's mental states. Since then, several hybrid views have been proposed. As well argued by Goldman (2006), simulation and theorizing need not be in competition. Rather, they may cooperate (if for instance, a theory is used to select the pretend inputs). Furthermore, some instantiations of mindreading may result from simulation only, whereas others may result from theorizing only. Nonetheless, even in its hybrid versions, at least one disagreement remains. In a nutshell, the priority of 1st person mindreading over 3rd person mindreading is at the core of Goldman's view, whereas most Theory theories and Rationality theories do not posit any asymmetry between 1st person and 3rd person mindreading (see for instance, Gopnik, 1993).

When analyzing the relation between 1st and 3rd person mindreading, one can ask two questions: (i) do they rely on the same psychological mechanisms or processes? (ii) if they rely on different mechanisms, are they fully independent or does one require the other? The debate is best illuminated if we compare Alvin Goldman and Peter Carruthers, who sit at two opposite sides of the debate. According to Goldman (1993a, 1993b, 2006), different processes are at stake in 1st person and 3rd person mindreading, with the latter depending on the former. Self-knowledge relies on a kind of inner sense or introspection, which utilizes an innate code in the language of thought, whose basic elements are caused by the various mental state types.

On the other hand, 3rd person mindreading consists in taking another person's perspective and in operating offline on the initial pretend states to generate further states, which can then be ascribed to the other person. Consequently, one needs to access one's own pretend states to be able to understand other people. There is a primacy of 1st person mindreading over 3rd person mindreading.[1] In contrast, according to Carruthers (2009), 1st and 3rd mindreading do not result from different mechanisms as far as propositional attitudes are concerned. Rather, they both involve interpretation. In particular, 1st person mindreading consists in swift unconscious self-interpretation. Hence, there is just a single faculty involved in both types of mindreading, using essentially the same type of inputs. There is no priority of self-knowledge of propositional attitudes.

Now a common complaint from the psychology side of the mindreading literature has been that one cannot empirically settle the debate between Simulation theory and its rivals (see for example Apperly, 2011). However, one may suggest that the analysis of mindreading biases offers a promising way to test the various theories, including the hypothesis of the (a)symmetry of 1st person and 3rd person mindreading. More particularly, according to Goldman (2006) and Goldman and Sebanz (2005), only the simulation theory can account for egocentric bias.

The clearest example of egocentric bias can be found in young children. In the classic version of the false-belief task, children observe the puppet Maxi putting a chocolate bar in the kitchen cupboard and leaving the room. Meanwhile Maxi's mother comes in the kitchen and puts the chocolate bar in the fridge. Then Maxi comes back. Children are asked where Maxi will search his chocolate bar when he comes back: in the cupboard or in the fridge. It has been repeatedly found that children under four years of age typically answer that Maxi will search in the fridge. They assume that Maxi has the same belief they have. In other words, they are victims of egocentric bias. But children are not the only ones to be sensitive to such bias. For example, Keysar and colleagues (2003) asked adult participants first to hide an object in a bag, unbeknown to the experimenter. Then the experimenter gave them rough descriptions of objects, which they were asked to move around in a grid. The description could sometimes better characterize the hidden object in the bag than a visible object in the grid. Yet, the correct response was to move the visible object, thus taking into account the experimenter's ignorance of the object in the bag. However, it was found that the participants

[1] However, simulation theories are not necessarily committed to first-person priority. For example, Gordon (1996), one of the earliest proponents of the Simulation theory, wonders about the legitimacy of the inference "from me to you" that is required by the simulation process as described by Goldman.

frequently took the object in the bag. There are many other examples of egocentric errors both in experimental conditions and in everyday life (for review, see Goldman, 2006, ch. 7 and Apperly, 2011, ch. 5). No matter how old we are, we can sometimes forget that other people do not share our beliefs and desires.

Egocentric errors have been taken as evidence of the priority of 1st person mindreading over 3rd person mindreading, which seems in line with the Simulation theory. When taking another's perspective, the subject can never fully pretend to be the other, that is, to have all the other's mental states. Rather, the offline use of psychological processes takes as inputs some pretend states but also some of the subject's own mental states (Nichols and Stich, 2003). This is not a problem if the other shares the subject's mental states, if they correspond to some commonsense assumptions for example. But this can lead to mindreading errors if the other has different beliefs or desires, as in the false-belief task. It is then especially important for the subject to quarantine her own mental states. Failure to do so leads to egocentric errors.

While one can easily account for egocentric errors within the simulationist framework, it seems more difficult to do so within the framework of the Theory theory and the Rationality theory, the two other main rivals to Goldman's view. If indeed 3rd person mindreading does not recruit 1st person mindreading, it is hard to see how the subject's own mental states could interfere with the understanding of other people. One possible answer could be that the folk psychology used by mindreaders includes a 'like-me' rule among other psychological laws. On this rule, the subject assumes that other people's mental states are like her own mental states. This may be plausible in children, although we shall see that even young infants do not always apply this rule. But Goldman argues that it is less intuitive in adults, and yet, they are also victims of egocentric errors even in simple situations in which one might expect them to make no error like in Keysar and colleagues.

One may not be as confident as Goldman that egocentric errors can be used as an argument in favor of the Simulation theory. On the one hand, some results interpreted in terms of egocentric mindreading errors may just reflect the *lack* of 3rd person mindreading. For example, one may suggest that in Keysar and coll. (2003), participants do not try to understand the experimenter's communicative intention about which object she has in mind; rather, they merely match a physical description to the object that fits best. As concluded by the authors themselves, "Under these circumstances, directly computing what another person knows or does not know at a given moment might be more trouble than it is worth" (Keysar et al., 2003, p. 39). If this interpretation is correct, these specific results are not relevant for 3rd

person mindreading because the participants do not use a mindreading strategy. In addition, even for egocentric errors that result from mindreading errors and cannot be reduced in such a way, it is not clear that the Simulation theory only can explain them. For example, Wallin (2011) interprets them within the framework of the Rationality theory. It is sometimes rational to attribute one's own mental states to other people. On this view, egocentric errors are merely a collateral damage of an efficient heuristic. Hence, it seems that one cannot use egocentric errors to settle the debate between the Simulation theory and its rivals. The Simulation theory, however, has to face a more important problem than the fact that other theories can account for egocentric errors. Indeed recent results seem to question the priority of first-person mindreading. As we shall see now, we can also make *altercentric* mindreading errors. But can the Simulation theory account for them?

2. When the other takes over

The landscape of the experimental study of mindreading has greatly changed these last few years. In particular, new versions of the classic false-belief task have appeared, showing that egocentric bias might be present only in explicit verbal behavior, but not in more implicit behavioral cues (for review, see Baillargeon et al., 2010). In new versions of the false-beliefs task, children are not asked any question about the puppet. Rather, the experimenter analyzes where the children look, that is, where they visually anticipate Maxi to search for the candy bar. It was found that infants in the second year of life rightly looked toward the cupboard. Alternatively, the experimenter tests whether children look longer when Maxi acts in a manner that is inconsistent with his false belief. Again, it was found that young infants look longer when Maxi looks for the chocolate bar in the fridge. Young infants may be less egocentric than expected. As we shall see now, the same may be true of adults as well.

How do you describe spatial relations between objects? The book is on the right of the bottle, will you spontaneously say, and this is so from your own perspective. However, a recent study showed that one could spontaneously switch to another individual's perspective, even when the situation did not require it (Tversky and Hard, 2009). When participants saw a photograph of a bottle and a book on a table, with a man seated behind the table, about a quarter of them described the spatial relations from the man's perspective (e.g., the book is on the left of the bottle in a framework centered on the man), although they had no interaction with the man depicted in the photograph. Along the same line, it was found that participants used different strategies to judge hand laterality when two participants facing each other

simultaneously performed the task and when only one was doing the task while the other closed her eyes (Böckler et al., 2011). The analysis of the reaction time relative to the angle of rotation showed that when alone, the participant used motor imagery to mentally rotate her own hand to match the displayed hand (egocentric frame). By contrast, in joint situation, the two participants mentally mapped the displayed hand onto body axes (allocentric frame). Yet, the fact that another individual was performing the same task at the same time had no relevance for the task itself. The participants themselves were not aware that the other's presence had affected their performance. Thus, the presence of other people, with whom one does not interact in any way and which should be normally irrelevant here, can lead to a switch of spatial reference frame.

Not only can we spontaneously take another person's visuo-spatial perspective, but we can also lose our own perspective in the process. This is so even when the other has a false belief, while we have a true belief, as illustrated by the following study. In Kovacs and coll. (2010), participants see a video showing a small story with a smurf, a ball and a screen. The story varies (e.g., the ball stays or not where the smurf has put it in the smurf's presence or absence; the ball reappears 'magically' at the end or not behind the screen, etc.). In one condition, the smurf comes back at the end and the screen is removed showing the ball. Participants are then asked to press a button as soon as they detect the ball. It was found that if the participants believed that there was a ball behind the screen, they replied more quickly than if they did not believe that there was a ball. Interestingly, participants were as quick to detect the ball (i) when they expected the ball and (ii) when the smurf expected the ball, although the participants themselves did not expect it.[2] Participants therefore took into account the smurf's belief, although it was in contradiction with their own true belief and it was not required by the task. Finally, participants replied more quickly even if the smurf was absent when the screen was removed.

A further study showed that another's beliefs could influence our judgment even when explicitly required to focus on our own perspective (Samson et al., 2010). Participants saw on a screen a room with an avatar and red discs displayed on one or two walls. In one condition, the avatar could see all the discs in the room. In another condition, the avatar could see only some of them (e.g., she turned her back to a wall with discs). Participants were then asked

[2] In this key condition, participants see the smurf hiding a ball behind a screen, the ball rolling out of the screen and the smurf leaving the room. They then see the ball going back behind the screen in the smurf's absence. Thus, the smurf falsely believes that the ball is still behind the screen and the participants rightly believe that there is no ball.

how many discs there were either from their own perspective or from the avatar's perspective. Interestingly, even when participants were explicitly asked to judge based on their own perspective, their answer was affected by the avatar's perspective. More precisely, they took significantly more time to respond and they made more errors.

In all these studies, another individual's perspective (or belief) affected the participants one way or the other, although it did not seem to be directly relevant for the task. One may draw two conclusions on the basis of these results, what we call the Automaticity claim and the Altercentricity claim. On the one hand, these results indicate that we mindread other people even when nothing seems to require it. One may then take these results as evidence in favor of the automaticity of some components of mindreading (Kovacs et al., 2010; Samson et al., 2010; Apperly, 2011). According to the *Automaticity claim*, we cannot help but mindreading people around us. 3rd On the other hand, these results indicate that another's perspective seems to dominate our own perspective in some circumstances. The output of third-person mindreading to some extent can unduly influence our mental processes and behavior. According to the *Altercentricity claim*, the other's perspective can intrude in our mental life. One may then be tempted to compare these results with the Stroop effect. Although irrelevant, the color the word refers to interferes with the correct naming of the color of the ink. Shall we then conclude like Tversky and Hard (2009, p. 129) that in some circumstances, "taking the other's perspective appears to be more natural and spontaneous than taking one's own"?

A few words of caution are necessary at this stage, as it is always the case if one wants to draw theoretical conclusions upon the basis of empirical results. First, it is worth noting that there are important differences among the studies described above. In particular, it is not clear that they all involve a mindreading component (e.g., Tversky and Hard, 2009). In addition, Böckler and coll. (2011) highlights the importance of joint attention, but it is questionable whether the other studies involve joint attention. Finally, one may want to distinguish between *altercentric bias* (adopting another's perspective when unnecessary) and *altercentric error* (adopting another's perspective when detrimental). When participants describe the relation between the book and the bottle, they display altercentric bias. When participants judge that they see only one disc rather than the two they can see from their own perspective, they make altercentric errors.

One may also want to question to what extent the other's presence (whether it is a man in a photograph, a virtual avatar or a smurf) cannot appear as relevant in these studies. It is true that the task (such as counting the discs in the virtual room) does not require taking into

account the other's perspective. But subjects are aware that they participate in an experiment, in which in general each detail matters. Hence, third-person mindreading may be motivated by the experimental context. To what extent do the results described above can then be taken as evidence of the automaticity of mindreading? It is all the more difficult to answer this question that the notion of automaticity is often left undefined. More interesting may be the claim that 3rd person mindreading is *mandatory*. Arguably, a process is mandatory if it is stimulus-driven (passive stimulation) and immune to interference. But the results so far do not suffice to show that 3rd person mindreading is purely stimulus-driven, independent of the context.

The Automaticity claim, however, is not at the core of our interest here. Rather, we are interested in the Altercentricity claim, and its implication for simulation-based accounts of mindreading. Even if the specific context of these studies can account for the apparently unnecessary 3rd person mindreading, it cannot explain the fact that to some extent the other's perspective or belief seems to take over the participants' own perspective or belief. One may, however, regret the weakness of the altercentric effects in these studies. For instance, Samson and coll. reported an increase in error rate due to altercentric bias. Participants who thought that an avatar thought there were, say, two discs in a display when in fact there were three were more likely to make mistakes than if the avatar's attributed beliefs were correct. It is important to note, however, that the difference in error rate was only observed in two out of their three experiments. In these two experiments, participants had to shift their responses, answering sometimes from their own perspective and sometimes from that of the avatar. Given that the participants were facing a speed-accuracy tradeoff (participants had to react as quickly as possible), it is not very surprising that they would sometimes make mistakes, such as taking the avatar's perspective when they had not been asked to do so on a specific question. By contrast, in the third experiment, participants never had to take the avatar's perspective. They then stop making mistakes. Moreover, while reaction times still increased when the avatar's perspective was inconsistent with the participants' perspective, the increase was small, and so hardly behaviorally significant. Interestingly, in the smurf study, the reaction time did not increase when the smurf's belief was inconsistent with the participants's belief. Better-controlled experiments are thus needed to make the altercentric bias more salient. Nonetheless, these preliminary results invite us to consider in more detail the notion of altercentric bias, in particular within the simulationist framework.

3. Altercentric errors in simulation

If egocentric errors cannot settle the debate between the Simulation theory and its rivals, then altercentric errors may work better. This time, however, the empirical evidence seems to run against Goldman's view. The fact that both egocentric bias and altercentric bias exist depending on the situation may indicate symmetry between 1st person and 3rd person, which is fully in line with Theory theories. Furthermore, altercentric errors seem to reflect a priority of 3rd person mindreading on 1st person mindreading in some circumstances. How is it compatible with the Simulation theory? Since Goldman in his hybrid view acknowledge that we sometimes use theorizing rather than simulating in our understanding of other people, he can always reply that altercentric effects occur only in these specific cases when we do not attempt to take another's perspective. Nonetheless, this reply seems slightly unsatisfactory and hardly compatible with Goldman's overall emphasis on the first-person. We shall propose here another answer for proponents of the Simulation theory, which offers an account of altercentric effects within the simulationist framework itself. Furthermore, we shall argue that far from being incompatible with altercentric bias, simulation opens the door to such bias.

It is first important to precisely understand what is at stake in altercentric errors. In particular, one should not assume that altercentric bias is the mere opposite of egocentric bias. They do not refer to one and the same phenomenon with a change of direction (from self to other in egocentric bias and from other to self in altercentric bias). Egocentric errors result from failure in 3rd person mindreading. The subject mistakenly ascribes to the target her own mental state. In contrast, in altercentric errors, the subject does correctly understand the target's perspective or belief. Rather, the mistake consists (i) to some extent in mindreading the target when not required by the task, which may appear as an unnecessary cognitive cost, but more importantly (ii) in being unduly influenced by the target, especially when the target has a false belief (cf. the smurf's belief that the ball is there in Kovacs et al.). In other words, whereas egocentric bias reveals under-use of 3rd person mindreading, altercentric bias reveals overuse of 3rd person mindreading. If now we compare egocentric and altercentric errors within the simulationist framework, we can see that they correspond to different stages in the simulative process. To recap, the simulative process can be articulated into four steps:

- Input selection: selection of the information both about the context and about the target that is relevant for the simulation.
- Offline use of psychological processes: emotional process, decision-making process, reasoning process and so forth are run off-line fed by the selected inputs.

- Output of psychological processes: the off-line use of psychological processes gives rise to emotions, intentions, beliefs and so forth.
- Ascription of the output to the target: Emotions, intentions, beliefs and so forth are then attributed to the target.

As previously argued, egocentric errors result from inappropriate input selection. Rather than pretending to have the target's beliefs and desires, one takes one's own beliefs and desires to feed the simulative process despite the fact that they are irrelevant. We will now argue that not only the 4-step model can account for altercentric errors, but it actually invites them. In particular, the third stage leaves open the possibility of confusing self and other.

Goldman (2006, p. 186) raises two questions about the simulative process: who is the subject of the pretend states? And what are the tags associated with them? He replies: "The mentalizer is the subject of those states (…) On the other hand, she can label, or tag, her pretend states as belonging to somebody else." What is important here is the fact that at stage 3, I am the subject of the output of the offline use of the relevant psychological process. For example, in the smurf study, when I put myself into the shoes of the smurf who sees the ball and then leaves, *I myself* entertain the (pretend) belief that the ball is behind the screen. Hence, I am the person who expects the ball to be there, even if I saw the ball rolling out of the screen. Altercentric errors can then be easily explained. There are actually two possible scenarios. First, the simulative process is disrupted. Arguably, we can conceive that the 'smurf tag' is left out for one reason or another (stage 4). That leaves open the possibility that I label the pretend belief as my own belief. It is then normal that this belief guides my behavior. However, the simulative process does not even need to be disturbed. Let us imagine now that the simulative process goes smoothly. My pretend belief is then associated with a smurf tag. But my pretend belief is erroneously taken as input to guide my behavior instead of my actual belief. Altercentric errors can thus result from errors either at the simulative level (wrong tag) or at the executive level (wrong input). In both scenarios, they reflect confusions between self and others: I take the smurf's beliefs as my own or I am acting on the basis of what the smurf believes rather than on what I myself believe.

These two scenarios are all the more plausible that they can occur in other domains. Let us start with the failure of tagging scenario, which is well illustrated by the example of emotional contagion, by contrast to empathy (for further details, see Vignemont & Jacob, forthcoming). In empathy, I share your anxiety, but I am aware that I am in this emotional state because you are anxious. I can then ascribe anxiety to you. In emotional contagion, on the other hand, I

share your anxiety, but I am not aware that you are at the origin of my feeling anxious. Rather, I catch your emotion, so to speak, and appropriate it as my own. In the conceptual framework we just proposed, I fail to tag my state of anxiety as yours. One may even suggest that emotional contagion is just a specific type of altercentric error.[3] Likewise, there are other instances of the executive scenario. For example, if you are a soldier, you may believe that it is a bad idea to attack during the night, but your officer believes it is the right thing to do. You then act on the basis of your officer's belief rather than your own. To some extent, obedience involves the same mechanisms as altercentric effects.[4]

To conclude, altercentric bias is fully compatible with the Simulation theory. The first person is so much at the core of 3rd person mindreading processes that one can actually forget that one is doing 3rd person mindreading. While this removes one possible difficulty for the Simulation theory, this also brings a new threat. If the Simulation theory does indeed invite confusion between self and other, one may wonder if the Simulation theory is desirable from an evolutionary perspective. One may suggest that the evolution should have selected a mindreading process that prevents as much as possible interpersonal influence, including altercentric errors.

4. Altercentric bias from an evolutionary perspective

A long philosophical tradition, marked most notably by Reid (1764) and Hume (1748), has tried to answer the following question: do we—or can we—naturally trust what other people tell us? Recently, psychologists have started to weight in. For instance, Daniel Gilbert has purportedly shown that "you can't not believe everything you read". People would first accept all communicated information before being able to question it in a second processing step (Gilbert et al., 1993). When the second processing cannot run its normal course, people may accept information that they would otherwise realize is false. While not identical, these findings and the altercentric bias share an interesting feature: information derived from someone else has an influence on us despite the availability of a negative assessment of that information. In this sense, both results suggest that considering information derived from

[3] Other examples may be found in the literature on action, and more particularly on the hypothesis of action mirroring (low-level simulation in Goldman's terms). Neuroscientific evidence indicates that some representations of action are recruited not only when we act, but also when we observe another agent acting. Furthermore, It has been argued that these motor representations shared between self and others can lead to confusion including self-attribution of other people's actions (for review, see Jeannerod, XXX)

[4] There are, however, differences. Obedience is – more or less – voluntary, whereas the allocentric effects we describe are not under voluntary control. Furthermore, there is a normativity in obedience that is absent in allocentric effects.

others as false and maintaining the separation with other beliefs deemed to be true are effortful tasks that can easily be derailed. These results thus support the view that trust towards information derived from others is a more natural stance than distrust.

About ten years ago, Dan Sperber (2001) shone an evolutionary light on this debate (fittingly enough, in a special issue of *Philosophical Topics* dedicated to the work of Alvin Goldman). In an evolutionary perspective, it does not make much sense to be naturally or overly trusting. Individuals' interests very rarely overlap perfectly. As a result, even when they have an incentive to cooperate and communicate, some component of competition is retained. In the case of cooperation, it means that cooperators are often better off "cheating:" reaping the fruits of cooperation while making the less effort possible. Cooperators have to be attuned to the existence of such cheaters if cooperation is to remain stable (Cosmides, 1989). Similarly speakers can influence listeners in ways that would only benefit the speakers and may be detrimental to the listeners—in other words, communication allows lying, cheating, manipulating. If communication is to remain stable, listeners have to be wary of misleading information communicated by speakers (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984). To put it simply, individuals who could too easily be led to accept wrong beliefs—death caps are edible, lions like to be petted—would not have passed on their gullibility to many descendants. Accordingly, it should not be a surprise that Gilbert et al.'s results—along with other instances of surreptitious influence such as subliminal persuasion—have been heavily criticized. Further experiments have shown that people can be lead to accept information they should reject only if participants have little previous knowledge on that topic and if the information is rather irrelevant (see, Mercier, in press-a).

Prima facie, the altercentric bias also lends support to the view that people can be unduly influenced by others. However, in the experiments reviewed above, there is no communication. Instead, participants are influenced by what evolutionary biologist would call a *cue*: the individual being observed (e.g. the smurf) is not trying to influence the observer in any way; it is the observer who does all the work (see, Scott-Phillips, 2008). Through communicative or non-communicative behavior, individuals could lead others to think that they entertain false beliefs. These false beliefs would then have an impact on the behavior of the 'victim' even if she knows they are false. The authors of the studies are aware of the dangers that can arise from altercentric bias: "The finding that others' beliefs can be similarly accessible as our own beliefs might seem problematic for an individual, because it may make one's behavior susceptible to others' beliefs that do not reliably reflect the current state of affairs" (Kovács et al., 2010, p.1834).

The problem is that we have seen that simulative processes open the door to such interpersonal influence by allowing confusion between self and other either at the mindreading level or at the executive level. From an evolutionary perspective, this danger could be so great as to shed doubt on the validity of the Simulation theory. As far as we know, Goldman has never directly addressed this problem. However, proponents of the Simulation theory may want to reply that altercentric errors are not more of a problem than egocentric errors. To avoid being victims of egocentric errors, one must quarantine's one's own beliefs and desires. Similarly, it may be suggested that to avoid being victims of altercentric errors, one must quarantine one's own *pretend* beliefs and desires. If well quarantined, one's own pretend beliefs and desires should not contaminate other cognitive domains. They should have no effect on executive control or decision-making. Most probably, quarantine plays a major role in preventing altercentric errors, but does it suffice from an evolutionary perspective? It seems that egocentric errors are quite frequent, which in simulationist terms implies that failure to quarantine is frequent. One should then expect altercentric errors to be frequent as well. This runs against the hypothesis that 'blind' trust is not advantageous from an evolutionary perspective. Interpersonal influence must be limited and under control. So we can ask: are we endowed with specific mechanisms designed to ward off manipulation attempts? Are they required to the same extent by communicative and non-communicative situations?

Most studies of the dangers of interpersonal influence have focused on linguistic interactions. Sperber and his colleagues have suggested that humans are endowed with a suite of mechanisms designed to ward off the dangers—mentioned above—raised by communicated information (Sperber et al., 2010). They point out that people should exert *epistemic vigilance* when they deal with communicated information. Thus, people adjust their trust according to the perceived benevolence and competence of the speaker (Mascaro & Sperber, 2009), they tend to reject information that conflicts with their previous beliefs (Mercier, in press-b) and they evaluate arguments aimed at persuading them (Mercier & Sperber, 2011). By contrast, such vigilance is hardly necessary when it comes to our own perceptual or inferential mechanisms, which were designed for our better good and proved to be remarkably reliable.

Given the preponderant role played by language in human interaction, the focus on linguistic communication is quite justified. Yet people can influence each other through other communicative means, even if they are not ostensive like language. In particular, the expression of emotion is a powerful mean of communication (ref?). Since most of the

information we derive from other people is acquired through communication, the dangers of being misinformed or manipulated through communication are also more important than those raised by non-communicative influence. Still, these dangers do exist. Such non-communicative influence can be passive—as when the observer attributes a belief to the smurf simply because he is gazing in a given direction—or active—for instance, the smurf could have been intentionally gazing in that direction so that the observer attributes to him a given belief. If people are not careful about the information they infer from other people's non-communicative behavior, they can easily be misled. A simple example will be used to illustrate the differences and commonalities between the treatment of information derived from communicative and non-communicative behavior. In particular, it will show that communicative behavior is more likely to change people's mind than non-communicative behavior.

Imagine there is a game in the newspaper with a trick question and a reward for the people who send in the correct answer. Paul's roommate, Lara, has written down an answer on her copy of the newspaper, but Paul thinks that the answer is something else. Paul then faces several options:

(i) He keeps believing that Lara believes that is the correct answer and either
   a. He changes his mind
   b. He believes she has made a mistake

(ii) He changes his mind about what Lara actually believes and either
   a. Attributes to Lara the intention to mislead him: she does not actually believe that is the correct answer but wants him to believe it.
   b. Changes his interpretation of Lara's behavior (Lara never thought that that was the correct answer, for instance she was just writing a possible answer on her way to figuring out the definitive answer.)

Now compare that situation with one in which Lara *tells* Paul: "I think the answer is X." Obviously, there are cases in which ostensive communication will be much more ambiguous than in the present example but overall the attributions that result from ostensive communication tend to be much less ambiguous than those resulting from the observation of non-communicative behaviors. Otherwise communication would be mostly moot. Option (iib) becomes much less likely in the case of communication, and the other interpretations are thus necessarily strengthened. There is no reason that (ib) or (iia) should be more likely in the case of communication than in the case of non-communicative behavior, and so Paul is more likely to change his mind when he has grounded his attribution in communicative behavior. This is

so not because he trusts Lara's competence or benevolence more, but simply because the interpretation is more ambiguous.

In addition, communication gives listeners an extra reason to change their mind because of the benevolence of the communicator: "By the very act of making an assertion, the communicator indicates that she is committing herself to providing the addressee with genuine information, and she intends his recognition of this commitment to give him a motive for accepting a content that he would not otherwise have sufficient reasons to accept" (Sperber et al. 2010, p. 366).

Let us imagine that Paul sees Lara carefully writing down her answer in the newspaper and leaving it on the table where he takes his breakfast, opened on the page of the game. Paul can now be quite sure that Lara wants him to believe that she believes that the answer is X. Still, Paul does not have the extra reason to accept X as the correct answer that he has when Lara tells him "I think the answer is X." For Lara's non-communicative behavior to provide as strong a reason as her communicative behavior to accept the intended belief, Lara should not be able to deny wanting to influence Paul, which should make her less likely to try to trick him. Furthermore, Paul must be aware of that, which provides him with a reason to accept Lara's belief. But how often does such a farfetched scenario (Paul has to know that Lara knows that he knows that she wants him to believe that she believes that the correct answer is X) happen? Other types of commitments can replace the intrinsic level of commitment found in communication. For instance, Lara could ask Paul to mail her answer to the newspaper (without saying what it is). If Paul looked up Lara's answer then, he would have a good reason to believe that it is actually what she believes to be the correct answer. The fact that extra proofs of commitment are necessary to make non-communicative behavior as credible or more credible than communicative behavior only demonstrates that communicative behavior is usually understood as naturally committing the speaker. Communication makes things much simpler (at least when it cannot be denied that communication took place at all, which is usually the case with verbal communication).

We have already seen two reasons why communicative behavior is more likely to influence other people than non-communicative behavior. A third reason could be an intrinsic suspicion of non-communicative metarepresentational intentions. When we think that someone intends us to think (or do) something, and yet the person does not rely on communication to achieve her end, we are entitled to doubt that her intentions are pure: otherwise, why would she not use the much more convenient way of transmitting information that is communication?

To conclude, altercentric effects point to an interesting but understudied phenomenon: the possibility of influence through non-communicative behavior. We have tried to show here that the mechanisms of epistemic vigilance that are used to deal with communicated information can also be recruited to treat influence through non-communicative behavior. Even if the same mechanisms are used in both the communicative and non-communicative cases, the dynamics differ, making non-communicative behavior less likely to successfully influence people than communicative behavior. In addition, there is a further difference between undue interpersonal influence in communication and altercentric errors, as it is illustrated in the smurf study for example. The difference is not only between communicative and non-communicative effects. The difference is also between intentional and non-intentional effects. Lara may be trying to influence Paul. But the smurf is not trying to influence the observer to his advantage. The subject is misled but the smurf is not responsible for it. If altercentric errors are merely the consequences of a failure to quarantine one's pretend states, then they can be seen as a rather innocuous computational bug, not more dangerous than when people make small mistakes in their understanding of the physical world. Furthermore, it is hard to see how individuals could come to make the best of the loophole in the quarantine process in order to exert undue influence on each other. Hence, it is true that the Simulation theory makes possible altercentric errors, which is not optimal from an evoluationary perspective, but it is not as bad as it could be because this cannot be used by others to manipulate the subject.

Conclusion

…

References

Apperly, 2011

Baillargeon et al., 2010

Böckler et al., 2011

Camerer et al., 1989

Carruthers (2009

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187-276.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (pp. 282-309). Oxford: Basil Blackwell Scientific Publications.

Goldman 1993a, 1993b

Goldman (2006)

Goldman and Sebanz (2005

Gopnik 1993

Gordon (1996),

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*(2), 221-233.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*.

Jeannerod, XXX

Keysar and colleagues (2003),

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830.

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* (Vol. 2, pp. 390-402). Oxford: Basil Blackwell Scientific Publications.

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

Mercier, H. (in press-a). Our pigheaded core: How we became smarter to be influenced by other people. In B. Calcott, R. Joyce, & K. Sterelny (Eds.), *Evolution, Cooperation, and Complexity*. Cambridge: MIT Press.

Mercier, H. (in press-b). The social functions of explicit coherence evaluation. *Mind & Society*.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.

Nichols and Stich, 2003

Reid, T. (1764). *Inquiry into the Human Mind*.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.

Scott-Phillips, T. C. (2008). Defining biological communication. *Journal of evolutionary biology*, *21*(2), 387–395.

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. (2010). Epistemic vigilance. *Mind and Language*, *25*(4), 359-393.

Tversky and Hard, 2009

Wallin (2011)