




The Social Cover View: a Non-epistemic Approach to Mindreading

Manuel Almagro Holgado¹ · Víctor Fernandez Castro² 

Received: 20 December 2018 / Revised: 20 May 2019 / Accepted: 27 May 2019 /

Published online: 17 June 2019

© Springer Nature B.V. 2019

Abstract

Mindreading capacity has been widely understood as the human ability to gain knowledge about the inner processes and states of others that bring about the behavior of these agents. This paper argues against this epistemic view of mindreading on the basis of different empirical studies in linguistics and social and developmental psychology: we are systematically biased in attributing mental states, and many everyday uses of mental ascription sentences do not reflect an epistemic function in our social interactions. We introduce an alternative view of mental ascriptions, the social cover view, which is consistent with the evidence. The social cover view holds that the main function of mental ascriptions is to cover the social status and reputation of an agent rather than to gain knowledge about her inner processes and states. Finally, we discuss two possible objections to our proposal.

Keywords Mindreading · Folk psychology · Social cover · Justification

1 Introduction

Making sense of others as minded creatures is a central competence of our social lives. Over the past three decades, philosophers and scientists have attempted to understand the cognitive processes underpinning the human capacity for understanding and interacting with others. Chiefly, this research has focused on so-called mindreading capacity, understood as the ability of humans to attribute or infer others' causal mental states in order to predict and explain their behavior. This research has been conducted on the premise that

✉ Víctor Fernandez Castro
vfernandezcastro@gmail.com

Manuel Almagro Holgado
manuelalmagroholgado@gmail.com

¹ University of Granada, Department of Philosophy I, Granada, Spain

² Institut Jean Nicod, CNRS UMR 8129, Département d'Etudes Cognitives, École Normale Supérieure & PSL Research University, Paris, France

human social interactions are an epistemic enterprise. Mindreading necessitates gaining knowledge about the inner processes and states which cause others' behavior in order to understand them. As such, our social understanding lies in the capacity to identify or gain access to the beliefs, intentions, emotions, desires and other mental states causing their behavior.

The aim of this paper is to challenge such epistemic view of mindreading on the basis of different empirical studies in linguistics and social and developmental psychology. Furthermore, we hold an alternative view about the function of mental ascriptions in our social interactions. This proposal, which we call the social cover view,¹ claims that the main function of mental ascriptions is not to gain knowledge about the inner states of others, but to cover or protect the social status and reputation of a particular agent; that is, to avoid public sanctions or potential conflicts. In what follows, we introduce the epistemic view of mindreading (section 2). In sections 3–6, we review several empirical studies that conflict with the epistemic view. In section 3, we review evidence regarding egocentric (Pronin et al. 2002) and in-group biases (Ferguson and Kelly 1964). In section 4, we present Malle et al.'s (2007) studies that suggest that subjects cite reason explanations and mental state attributions more often when they are the actors or when they want to portray someone in a positive light. In sections 5 and 6, we present linguistic and psychological evidence suggesting the existence of non-epistemic functions of mental ascriptions, i.e., parenthetical uses (Aijmer 1997; Goddard 2003; Urmson 1952; Wierzbicka 2006) and the oblique uses of third-person mental ascriptions (Hansen et al. 2017). This evidence, we conclude, points against the idea that in attributing mental states we aim to know about the internal processes and causal states of agents. In section 7, we outline our proposal, and finally, two possible objections are discussed in section 8.

2 The Epistemic View

According to a widely held view in cognitive sciences and philosophy of mind, human socio-cognitive capacities mostly rely on a mindreading faculty, the capacity for understanding other creatures' behavior in terms of mental states. Generally, mindreading is construed as a mental device that infers or simulates others' mental states as a means to predict and explain their behavior.² Here is a standard characterization of how mindreading works:

¹ To our knowledge, the label "social cover" in the way we use it here was coined by Norman (2016). This label refers to the idea that reasoning mechanisms are used for projecting a positive social image rather than for epistemic purposes. Given that our proposal can be seen as an application of the same idea to mental states attributions, we have decided to keep this label.

² Traditionally, there are two groups of theories regarding mindreading. According to the theory-theory, mental state attributions are the result of implicit theorization based on a systematic corpus of general laws specifying the connection between perceptual inputs, internal states and behavioral outputs (Baron-Cohen 1995; Gopnik and Meltzoff 1997; Gopnik & Wellman 1994). According to the simulation-theory, the mindreading process is carried out by different simulation mechanisms based on introspections or off-line sub-personal mechanisms (Goldman 1989; Gordon 1996; Heal, 1998) that exploit the access to our own mental states to project what others would do in a given situation. In later developments, several authors have tried to develop certain hybrid versions involving some combination of the processes. In fact, nowadays there is a common consensus about the existence of both types of mechanisms (Carruthers 2006; Goldman 2006; Nichols & Stich 2003). For the purpose of this paper, we consider both approaches as instances of the epistemic view. Both views presuppose that laypeople tackle social encounters as instances of an epistemic problem.

[The] mindreading system containing concepts of *belief*, *desire*, and *see*. This system is engaged whenever an agent is identified as such, and it automatically tracks what the agent can see, using cues such as eye-direction and the saliency of objects and events in the agent's line of sight. These attributions, in turn, automatically give rise to attributions of the corresponding beliefs... In effect, we can suggest that the mindreading system automatically builds a partial model of the mental states of any agent that it encounters (Carruthers 2017, p. 682).

Humans exploit an array of general and contextual information which surrounds behavior in order to infer the internal states of other agents, to generate different predictions and explanations of subsequent behaviors. Notice that the rationale behind the idea is that, given that we only have access to crude behavior, human socio-cognitive capacities may be based on a mechanism or ability that reflects the inner processes and states that bring about the behavior (see for instance, Apperly (2011, p. 5), Bartsch and Wellman (1995, pp. 4-5, p. 115), Fodor (1992), Gopnik (1996, p. 187), Gopnik and Meltzoff (1997, pp. 13-42)). In this view, mental states ascriptions aim at predicting and explaining “the behavior of another agent by postulating that unobservable inner states particular to the cognitive perspective of that agent causally modulate that agent's behavior” (Penn & Povinelli 2007, p. 394). Without this ability, humans could not predict and explain each other's behavior and, therefore, they would be unable to coordinate and carry out social interactions.

In this sense, although in principle social creatures could ascribe mental states for different purposes, the standard explanation works within the view that the main function of mindreading is to gain knowledge about the non-observable causal states of individuals which bring about a particular pattern of behavior. In this view, the function of mindreading is *epistemic*. Zawidzki presents the view as follows:

Our epistemic relations to each other are no different from our epistemic relations to the nonhuman world. Other people are mysterious phenomena animated by unobservable causes that are completely independent of our attempts to understand them. Individual human beings must learn to infer these causes to better predict and control other human beings, just as is the case with nonhuman phenomena. They do so by attempting to represent these unobservable causes, and they succeed to the extent that they represent them accurately... In short, according to the mindreading metaphor, distinctively human social cognition is conceptualized as an individual accomplishment, involving the accurate representation of independently constituted, unobservable mental causes of behavior. (2013, pp. xii-xiii)

When we claim that the function of attributing mental states is epistemic according to this view, we do not claim that this view suggests that laypeople's central goal is to know or specify the internal causal processes of behavior. Instead, we believe that such an epistemic stance is assumed in their central thesis that we treat others as an object of prediction and explanation. As Meltzoff et al. put it:

In making sense of one another, we need to bridge a gulf between what we can ‘directly’ experience about other people, and what is going on ‘in’ their minds:

‘our sensory experience of other people tells us about their movements in space but does not tell directly about their mental states’. (Meltzoff et al. 1999, p. 17)

Thus, as several authors have emphasized (Gallagher and Zahavi 2008, p. 183; Zahavi 2011, pp. 547–548), the standard characterization of mindreading embraces the assumption that people face social interactions as being an epistemic endeavor. In particular, such characterization presupposes that everyday social interactions must be methodologically encountered as a counterpart of the classical epistemological problem of the other minds.

Against such an assumption, in the following sections, we review some studies that call into question the status of mindreading as an epistemic competence. Certainly, we do not deny that social creatures use mental ascriptions for epistemic ambitions on certain occasions. Our main claim is, nevertheless, that such function is not paradigmatic.³ As a result, we argue that the evidence supports the idea that ascriptions, more often than not, serve for reputation management or projecting a rational image of ourselves —i.e., providing reasons and justification for action, giving indirect evidence of our claims or signaling uncertainty that protects us from the social consequences of providing false information.

3 The Problem of Accuracy

This section aims to present several studies that offer reasons for doubting the epistemic view of mindreading. To see how, notice that, in the epistemic view, the idea that mindreading aims to gain insight into other minds assumes that given our efficacy in social coordination and anticipation, we must possess a reliable and efficient method for accessing such states. Our proficiency in predicting and explaining others’ behavior requires the identification of the inner mental states which bring about their actions. Thus, the epistemic view relies on the assumption that our epistemic access to others’ minds must be sufficiently reliable and accurate to generate stable predictions and explanations.

However, while our capacity for coordination in social contexts is overwhelmingly efficient, a number of findings in social psychology have demonstrated that humans are systematically prone to error in attributing mental states, due principally to bias, prejudices and cognitive limitations. For instance, it seems very common to classify people according to their social categories, like age, race and gender (Ito et al. 2004; Liu et al. 2002). These categories serve as a basis to rapidly attribute personality traits such as competence, dominance, aggressiveness and trustworthiness to others (Olivola and Todorov 2010; Rule et al. 2009), and to make implicit associations between these

³ The central claim of this paper is fully compatible with the pluralist views of social cognition, according to which socio-cognitive skills in general, and mindreading in particular, can exhibit different functions (see Andrews 2012; Fiebich and Coltheart 2015). However, different pluralist views can propose different/various specific hypotheses regarding the role of mindreading and social skills. In this respect, we believe that the arguments and evidence presented in this paper are sufficient to claim that the epistemic function is less frequent in everyday interactions than it has been assumed. Furthermore, as the evidence in developmental psychology suggested in section 5 and 6 indicates, we put forward the idea that the social cover function is developmentally prior to other functions (for a similar point see Fernández Castro 2019).

social categories and other specific characteristics, some of them measured by the Implicit Association Task (Greenwald et al. 1998; Greenwald et al. 2009, which have an effect on our mental attributions.⁴ The influence of these biases is not restricted to conditions where we are under cognitive load (Gilbert et al. 1988), such limitations in mental ascriptions also confine us in scenarios where we attempt to display more deliberative and careful ascriptions (Wittenbrink et al. 2001). In addition, self-interested purposes (e.g., anxiety reduction, confirmation of our worldview) also elicit errors in mindreading (Dunning 1999; Kunda 1990). Let us introduce two studies involving egocentric and in-group biases to illustrate this point.

In an empirical study conducted by Pronin et al. (2002), 24 undergraduate students were asked to rate, on a scale of 1 to 9, how much they, and the average American, showed eight different specific biases. Thirteen students were asked first about their own susceptibility to each bias, and then about the susceptibility of the average American. The rest of the students rated the average American before themselves. The results showed that participants perceive themselves as less susceptible to each of the eight biases than the average American. This is a common egocentric bias. The results remained relatively constant in other versions of the survey. Regarding this study, one may object that it could be the case that the undergraduate students are actually correct about being less biased than the average American. However, to think that one is immune to the effects of the experiment to which one responds is quite common. As Pronin herself shows through different versions of the study (Pronin et al. 2002; see also Pronin 2007), there is a blind spot: the tendency to represent oneself as uniquely immune to flawed reasoning. To rule out the alternative hypothesis that participants of the study, as students from a prestigious university, are making a comparison between their analytical skills and those of the average American, and to test again the idea that people tend to perceive themselves as less biased than other people, Pronin and her colleagues carried out two versions of the survey. In the first version, 30 students from the same psychology seminar course were asked to assess how susceptible they were to various biases in comparison with the rest of the seminar peers. The results of this second survey replicated the results of the first: the participants perceive themselves as less biased than the rest of their peers (Pronin et al. 2002: 371–372). In order to eliminate the possibility that the results were influenced by the fact that the participants are students from prestigious universities, they conducted another version of the survey. In the second version of the survey, 76 people of varied age and background who were waiting for their flights at an airport were asked to respond how widespread certain biases were among the travelers who were at the airport that day, and then how susceptible they were themselves to the same biases. The results of this third survey replicated the results of the previous two (Pronin et al. 2002: 372–374). People tend to think that they are less biased than other people.⁵ In another empirical study conducted by Ferguson and Kelly (1964), two groups of three to six

⁴ Although character traits like aggressiveness and trustworthiness can be considered mental states, we follow the standard literature which often distinguishes between occurrent mental states (e.g., beliefs, desires, intentions) and other attributions (e.g., traits, stereotypes). However, it is an open question whether or not the latter must be considered as part of our mindreading capacities or just an influence (see Spaulding 2018, pp. 24–36; Westra 2017b).

⁵ See also Hannikainen 2018 (p. 5) for a recent example where participants think they are immune to the effects of the experiment in which they participate.

members worked independently on three tasks. Both groups were asked to separately rate the quality of the products of both groups on a scale of 1 to 9. The results obtained showed that they were all biased towards evaluating the product of their own group more positively than the other group's product.

The results of these studies speak in favor of the idea that social and egocentric stereotypes, cognitive load and other contextual factors systematically condition our mental state ascriptions—to the extent that all these phenomena provide information that we use to attribute mental states—and make them prone to error.⁶ As such, our capacity to ascribe mental states is not as reliable and accurate as the epistemic view supposes. Accuracy is not only a problem for prediction, as we have already said. If explaining a behavior successfully has to do with knowing the internal processes of the subject, accuracy is also a problem for explanation. We could explain someone's behavior by virtue of the fact that we biasedly and erroneously attribute to her a mental state. This opens the possibility that the proficiency we exhibit in coordination and anticipation might not rely on mindreading (Andrews 2012; Mameli 2001; Zawidzki 2013); but most importantly, it calls into question the claim that the main function of mindreading is to gain epistemic access to other minds. Of course, someone may object that the core function of mindreading is indeed epistemic and that we are simply very bad at performing this function. What has been said thus far is compatible with this statement. However, the fact that we are systematically biased, along with the evidence we will discuss in the following sections, reasonably weakens the idea that the capacity to identify or gain access to other psychological states underpins our abilities for explanation and prediction.

4 The Problem of Explaining to Others

Accurate or not, the fact is that mental ascription is pervasive in our understanding of intentional actions. Therefore, testing the epistemic view requires us to pay closer attention to how mindreaders explain intentional actions.⁷ The rationale behind the epistemic view is the idea that our mental ascriptions serve to gain a better

⁶ One may recognize a different role for implicit biases in social interpretation while resisting to abandon the epistemic assumption (Spaulding 2016; Westra 2017b). For instance, Spaulding (2016) argues that social situations impose different needs for accuracy and efficiency, forcing agents to rely on shortcomings and heuristics when the situation demands efficiency, and rely on more reflexive and deliberative processes when accuracy is required. There are two problems with this view. First, it is not so clear that deliberative and reflexive processes are unbiased (Wittenbrink et al. 2001), which would cast into question even a more restricted version of the epistemic view, where accuracy is only needed in some situations. Second, some important evidence supports the idea that stereotypes and implicit bias are not as inaccurate as it is often assumed (see Jussim 2012 for a review), reinforcing the idea that the efficiency of our anticipatory and interpretative capacities may rely on different processes such as stereotypes and character traits, rather than mindreading (Andrews 2012, 2015; Hirschfeld 2013; Kalish 2012; Maibom 2007; Zawidzki 2013; section 4). Of course, taking into consideration this evidence goes beyond the scope of this paper. For our purpose, it is enough to notice that our mindreading capacities are prone to error in a way that accuracy is far from being secured.

⁷ Philosophical literature regarding action often distinguishes between intentional actions and mere happenings or passive behavior (see Wilson and Shpall 2016 for a review). For the purpose of our paper, we do not focus on such distinction; instead, we focus on those actions and behavior that the agent is accountable for. In this sense, we use the terms intentional actions and behavior interchangeably.

understanding of others' behavioral causes, which gives us a better understanding of others and could help us to make our social interactions more fluid. Given that, the epistemic view generates the prediction that in the right social situations, mindreading will tend to provide explanations in terms of internal causal processes rather than appealing to other types of explanations like enabling situational factors, traits or causal history of reason explanations. However, the empirical evidence in social psychology which we review in this section does not support this prediction.

Malle and his collaborators have carried out an intensive investigation aimed at exploring the different ways people explain actions (2007). Malle and his group distinguished between different modes of explanations (see for instance, Malle (2011, p. 308)). First, reason explanations, which consider an agent's reasons for intending to act, which are sustained implicitly or explicitly in mental state attribution (e.g., 'Anthony is running because [he thinks] he has been gaining weight'.)⁸ Second, causal history of reason explanations which appeal to external causal factors in the history of the agent rather than reasons that motivate actions (e.g., 'She told me to stay away from the neighbors' kids because we lived next to this really run-down apartment building and she was a bit overprotective.'). Third, enabling factor explanations which cite conditions that could facilitate the action as intended, but they are not in themselves reasons (e.g., 'She worked through the night because she had a lot of coffee.'). Given this framework, Malle and his group asked subjects to explain different actions, to test the type of explanation they would give depending on whether they were actors or observers of the action. The interest of these studies lies in the fact that people exhibit asymmetries regarding the type of explanation depending on whether they are the actors or the observers of the action. In particular, they found that people cite reason explanations (both marked and unmarked with mental states) many more times when they are the actors than when they are the observers, in comparison with causal history or enabling factor explanations.

These findings seem to be inconsistent with the idea that we are motivated to ascribe mental states to gain a better understanding of others. Of course, one may argue that the epistemic claim emphasizes the access to mental states in general and, naturally, agents might find it harder to infer others' mental states than their own. As Malle emphasizes, "for observers, information access is more challenging: they must rely on stored knowledge, inferences, and simulations, with the real possibility that no explanation can be found" (2011, p. 323). However, this explanation is not as plausible as it may seem at first glance. First, it is hard to see why, given the effortful nature of the process, people do not opt to abandon reason explanations entirely in favor of the other types of explanations, especially when a mindreader could exploit other types of information available that do not require epistemic access to mental states; for instance, biological motion and contextual factors (Ruffman 2014), social models (Maibom 2007), social norms (Hirschfeld 2013; Kalish 2012; Fernández Castro and Heras-Escribano 2019), stereotypes (Mameli 2001) or generalizations (Millikan 2004; Ruffman and Perner 2005).

⁸ Although Malle and his colleagues claim that reasons explanation and reasons explanation involving mental vocabulary are mostly the same (they involve psychological causes of the action), there is an important tradition in philosophy that denies that reason explanations are implicit psychological causes. According to these authors, reasons must be construed as facts or propositions that motivate the action (Alvarez 2010; Scanlon 1998). See below.

Second, there is an alternative explanation of the asymmetry that sounds more persuasive in the light of other findings in Malle et al.'s studies (2007). In one of their studies, they demonstrated that a way to reverse the observers' tendency to provide more causal historical factors than reasons was to motivate them to portray the actor in a positive light. In other words, when the subjects were instructed to make the agent look good, they systematically used more explanations in terms of marked and unmarked reasons. These results, when taken together with reason asymmetry, suggest that reason explanation in general, and mental attribution in particular, are aimed at putting the actor in a positive light. So, actors use more reasons and belief ascriptions because they are inclined to depict themselves or others in a favorable way. Furthermore, people offer many more reasons in terms of mental states when they face puzzling and unexpected actions (Korman and Malle 2016; see also Uttich and Lombrozo 2010) which seem to be circumstances where the attributer might feel forced to portray an actor's behavior as reasonable or intelligible. Thus, it is reasonable to think that these studies show that mindreaders provide explanations in terms of reasons or mental ascriptions when they are motivated to put the target's action in a positive light, and not to identify their mental state.

The upshot of these studies thus far is that people use ascriptions for providing a positive impression of the attributee's actions. In principle, such results are not incompatible with the central epistemic claim that ascriptions identify or reveal others' inner psychological states. Indeed, one may argue that the positive evaluations of mental ascriptions reflect the fact that we tend to put more effort into understanding and epistemically accessing those we like (e.g., our friends). In this view, we provide more mentalistic explanations whenever we feel well-disposed towards someone⁹ (Spaulding 2018, pp. 15–19). However, this does not mean that our ascriptions are not aimed at identifying the causal states bringing about their behavior.

Malle himself seems to motivate such interpretation when considering that marked and unmarked reason explanations refer to psychological causes. However, there are at least two reasons to doubt that impression management is compatible with the epistemic claim that ascriptions reveal inner psychological states. First, as Tanney (2013) has argued, people often avoid accepting an explanation for normative reasons even when it is epistemically accurate. Tanney (2013) exemplifies that in the following manner. Suppose that a firefighter is ready to run into a burning building to save some people trapped in it but, in the last moment, she decides to run out the building. Now, imagine we ask the firefighter why she ran out. She answers that she was starving and wanted something to eat, and so she left the building. Consider how awkward we would find this answer. As Tanney put it:

We would reject this as an explanation on the grounds that it does not make sense. Suppose she says that it makes perfect sense to her why she would drop her everything—even put lives at risk—because she wanted something to eat. We just do not understand what it is like for her when she wants something. Indeed, she is right: we do not understand. (Tanney 2013, pp. 143–144)

⁹ Thanks to an anonymous referee for bringing our attention to this.

The reason why such an explanation would not make sense to us is that according to our standards of rationality or acceptability, saving persons is extremely valuable. In this sense, even if the explanation is epistemically accurate, we would find the explanation inadmissible for normative reasons. In fact, we would tend to think that person is having an abnormal episode because the commitments acquired when deciding to prepare and enter to save the people trapped in the building are rationally incompatible with suddenly going to eat something. Thus, we evaluate reasons and ascriptions supporting actions in terms of norms of social permissibility or rationality, rather than in terms of epistemic accuracy regarding the real causal state that brings about actions.

Second, there is indirect evidence that indicates that we evaluate mental states ascriptions in rational normative rather than epistemic terms. In particular, such an idea finds support in different findings regarding the phenomena of confabulation and biased-reasoning. On the one hand, as Bergamaschi Ganapini (2019) has persuasively argued, the idea that reasons are psychological causes is hard to make compatible with how patients who suffer from autobiographical memory impairment provide reasons. For instance, a patient reported having been married for about four months when he had been married for 30 years (Moscovitch 1995). When asked about their children, he explained that they were adopted (which was not true). The reason ‘my children were adopted’ cannot have played a role in the formation of his belief that he was married four months ago. Instead, such a reason is “used to fill a gap in the agent’s view of the world” (Bergamaschi Ganapini 2019, p. 6), and it does not seem to have anything to do with the psychological cause of the attitude. Furthermore, understanding Malle et al.’s findings in non-psychological reasons is coherent with the fact that our reasoning capacities, in general, seem to be directed towards persuading and convincing rather than aiming for the truth. Well-established literature in experimental psychology demonstrates that human reasoning is systematically biased (see Kahneman 2011 for a review). For instance, we are prone to provide reasons that reinforce our previous beliefs even when those beliefs are false (Nickerson 1988). Mercier and Sperber (2011, 2017, see also Norman 2016) have argued that this evidence suggests that human capacity for providing reasons did not evolve for deriving reliable information or helping us to make better decisions. Instead, “reasoning can lead to poor outcomes not because humans are bad at it but because they systematically look for arguments to justify their beliefs or their actions” (p. 72). In this view, we evaluate and provide reasons intended to persuade others, and thus, our competence to give and evaluate reasons is associated with justifying beliefs and behaviors in group contexts (Mercier and Sperber 2011, pp. 62–63); but also, with supporting and protecting the members of our group (Brewer and Brown 1998; Pettigrew 1979).

In a nutshell, these findings regarding confabulation and biases in reasoning, along with Tannev’s theoretical argument regarding normativity, seem to suggest that mental states explanations are non-causal explanations. Furthermore, these findings and arguments suggest a robust framework to understand Malle’s findings where mental ascriptions can be understood as aimed at rationalizing or normalizing actions for social cover purposes. Mental states attributions are regarded as explanations that contextualize the action by indicating contextual factors that rationalize or normalize

one's action (Tanney 2013; Sanchez-Curry 2018; Strijbos and De Bruin 2011), and thus, they project a positive image of ourselves as normative rational agents (see Bergamaschi Ganapini 2019, p. 1).

5 The Parenthetical Function

In the previous section, we have claimed that some relevant literature in social psychology conflicts with the idea that mental ascriptions are aimed at gaining epistemic access to others' mental states. Problematically for the epistemic view, there are other empirical findings that indicate other uses of mental ascriptions that are difficult to explain within its framework. In this section and section 6, we review empirical studies suggesting the existence of non-epistemic functions of ascriptions. These uses are not only pervasive in our everyday interactions but also seem difficult to account for, when considering only the epistemic view.

Several developmental psychologists have reported that the first use of mental verbs by children is non-epistemic (Bartsch and Wellman 1995; Sabbagh and Callanan 1998; Shatz et al. 1983). Instead, they exhibit some uses of mental ascriptions that serve to pursue a contrastive or pragmatic function, that is, to modulate the assertion or proposition attached to the mental verb. For instance, the phrase 'I think' is frequently presented as having the function of mitigating the degree of commitment to the sentence it accompanies (e.g., 'it's raining, I think'). These contrastive uses of mental predicates are what linguists and philosophers have called parenthetical uses (Aijmer 1997; Goddard 2003; Urmson 1952; Wierzbicka 98). From a syntactical point of view, parenthetical verbs are verbs that appear in the first-person present in any of these forms: 'I suppose (that) your house is very old'; 'Your house is, I suppose, very old'; 'Your house is very old, I suppose'. From a pragmatic point of view, these uses of mental verbs serve to indicate how the statement accompanied by the verbs must be interpreted. Wierzbicka (2006) provides a deep analysis of parenthetical uses of 'believe', 'think' and other mental verbs. She claims that the verb 'think' conveys the meaning of disclaiming knowledge not by saying 'I don't know' but by saying 'I don't say: I know'.

Given the centrality that philosophers have given to the idea that ascriptions are epistemic tools, they have often overlooked this type of pragmatic use of mental verbs. In fact, one may think that such uses are non-genuine uses of mental verbs.¹⁰ However, these uses are quite pervasive in adults' social exchanges. In a study based on audiotapes of informal conversations (Santa Barbara Corpus of Spoken American

¹⁰ It may be objected that, while it is obvious that mental state verbs are being used here, it is unclear that this use says anything about mental states concepts. However, it should be noted that the motivation for the mindreading research in psychology and philosophy was to offer an explanation of how human beings use mental attributions, and therefore how we use them in our linguistic exchanges (see for instance, Perner 1991, pp. 1–11; Stone and Davies 1996, pp. 119–120). If much of the mental vocabulary that we use is not meant to talk about the internal processes of agents, perhaps this is not what we mostly do by attributing mental states. To think that these uses of mental vocabulary do not really refer to mental concepts is to excessively restrict the 'pure' phenomenon of mental attributions, which in turn supports our idea that it is not the main function of what we do when we talk about people's beliefs, desires and other mental states.

English), Scheibman (2001, pp. 70–72) found that cognitive verbs (e.g., ‘to think’, ‘to believe’, ‘to know’) appear in first-person sentences far more often than in third-person sentences. Specifically, cognitive verbs, among which ‘think’ and ‘know’ stand out significantly, appear in the first person 195 times out of 340 (57%), while they only appear 15 times in the third person (4.4%). In a similar study, Thompson and Mulac (1991, pp. 320–222) showed that epistemic verbs occur in the first person with higher frequency than in the third person (95% vs. 1%) when they appear without the *that*, which in the case of *think*, for instance, is 91% of instances. This number of occurrences of epistemic verbs in the first person without *that* suggests that an important part of the uses of our cognitive verbs are parenthetical.¹¹ As Scheibman (2001, pp. 70–71) emphasizes, 89% of all present tense verbs seem to be formulaic expressions to mitigate assertions. This high frequency of parenthetical verbs in spoken language has also been reported by other linguists (Aijmer 1997; Goddard 2003; Thompson and Mulac 1991; Urmson 1952; Wierzbicka 2006).

These types of uses of mental states seem to be a default interpretation of our mental verbs. As the studies of Lewis et al. (2017) have shown, parenthetical uses of mental states could play an important role in the explanation of why children are not able to exhibit adult-like interpretations of belief ascriptions before their fourth year. In these studies, two scenarios were presented to several children using action figures. In the first one, the character Dora is out of a room while another character, Swiper, hides behind a curtain, leaving his yellow tail visible, and a third character, Squirrel, enters the room and hides behind a box to throw off Dora, also leaving his yellow tail visible. Dora comes back and looks for Swiper behind the toy box. Then, a puppet says “Dora thinks that Swiper is behind the toy box”. In the second scenario, a second seeker (Boots) was introduced and looks for Swiper behind the curtain. So, the puppet reported that while Dora mistakenly thinks that Swiper is behind the box, Boots correctly thinks that Swiper is behind the curtain. Children were asked to judge whether the puppet was right or not about what happened in both scenarios. The results of the study demonstrated that children interpreted the attributions in an adult-like manner (correctly attributing a false belief when appropriate) in the second condition but not in the first condition. This demonstrates that 3-year-old children can correctly attribute false beliefs when such an interpretation of the report is more salient (conflict beliefs). However, their default interpretation is to consider the report as expressing evidence; probably because they are more systematically exposed to this type of use. In other words, given the predominant use of *think* as endorsement in adults, a child may first acquire this use of the verb and misinterpret the report: she may misunderstand the report as citing Dora’s behavior as a source of evidence rather than as appealing to her belief (Lewis et al. 2017, pp. 5–6). The results support this hypothesis. This pragmatic explanation has also recently been offered by Dudley: a child can understand a sentence like ‘the agent thinks that the object is in location A’ as ‘the object is in location A, because the agent thinks so’ due to the ‘assertive’ nature of ‘think’ in everyday conversations that she is exposed to (Dudley 2017). Furthermore, as Dudley (2018) emphasizes, this explanation is consistent with other recent evidence showing that children can attribute false beliefs before the age of 4, for instance, when the subjects

¹¹ For a similar point regarding German and Dutch see Nuyts (2001, pp. 107–167); for the case of French see Schneider and And Glikman (2015).

are asked open-ended questions, and it is ensured that they can keep track of the attributee's perspective (see Rubio-Fernandez and Geurts 2012).

Now, these parenthetical uses of mental verbs do not exhibit an epistemic function (identifying the mental states of the speaker). Instead, the speaker specifies to the hearer how she must understand the proposition or assertion; for instance, signaling that she is not fully committed to the claim. Notice that indicating to the audience the degree of reliability in one's claim can protect oneself from possible accusations derived from the claim or reinforce one's confidence in it. However, this social function of mental verbs does not seem to require the identification or description of the mental state of the speaker, but the expression of some degree of truthfulness or reliability regarding the claim.

6 Ascriptions as Indirect Evidence

First-person ascriptions might not be so compelling as an argument against a theory that attempts to elucidate the role that mental ascriptions play in the interpretation of others. However, the parenthetical uses of mental verbs presented above are closely connected to a different non-epistemic, oblique use of third-person mental ascription sentences. In particular, the use of sentences with non-factive embedding verbs¹² (e.g., 'Kautar believes that the library is closed') are often interpreted as offering indirect evidence (e.g., Hazlett 2010; Kiparsky and Kiparsky 1970; Levinson 2000; Simons 2007; Urmson 1952). In these ascriptions, the speaker aims to provide support to the proposition expressed in the that-clause sentence by stating that a third person also supports it. Simons (2007) argues that this interpretation of belief attributions is the most plausible in situations like the following.

A: 'Is the library open?'

B: 'I don't know for sure. Kautar believes that it's open'.

In this case, speaker B seems not to be talking about Kautar's belief, but rather about how things are, that is, she is expressing a certain commitment (albeit only weakly) to the truth of the proposition *the library is open*.

Recently, Hansen et al. (2017) have conducted a research that supports Simons' claim. In the study, the subjects were presented with two contexts. In the ordinary context, speaker A asks B something like this 'I want the book. Where does Kautar believe it is?', and B answers 'On the table'. On the other hand, in the conditional context, speaker A asks B something like this 'Even though I have already found the book, can you nevertheless inform me where Kautar believes it is?', and B answers 'On the table'. Then, the subjects had to answer a questionnaire to test their interpretation. The experimenters found that most of the participants interpreted the belief ascription in the ordinary context as an indirect way to offer evidence about something in the world instead of a psychological description of that person's mind. Furthermore, belief

¹² Non-factive verbs are verbs used to show that the speaker is not completely committed to the truth of the proposition expressed in the that-clause.

attribution vignettes were contrasted with similar cases in which the attribution does not contain belief verbs, but belief-dependent verbs. Belief-dependent verbs are action-verbs which are used in situations in which behavior is motivated by a belief, like *look for* and *says that* verbs. Although both verbs refer to a behavior that depends on a belief, they do not refer directly to mental states and are not usually used to speak about people's minds. Interestingly, when they were presented in the ordinary context, all three verbs elicited an evidential interpretation from the participants.

The fact that we often use mental state verbs as an indirect way to offer evidence about something in the world instead of a psychological description of that person's mind seems to suggest that we use the attribution to provide support to our own claims. For instance, we exercise the other person's authority in order to strengthen the truth of our claim (see Van Cleave and Gauker 2010, pp. 314); but also, in other circumstances, we can use such indirect reports to avoid our own responsibility regarding the claim. However, what such uses have in common is precisely that they are used as tools for social management, rather than to give epistemic cues concerning the attributees' mind.

In summary, the review of different findings in linguistics and social and developmental psychology indicates that our everyday uses of mental ascriptions reveal different functions which are difficult to explain from the epistemic view. In the next section, we develop an alternative view that accounts for such functions.

7 The Social Cover

The reviewed evidence supports the idea that certain uses of mental ascription sentences do not reflect an epistemic function. Instead, we attribute mental states to others when (1) we want to portray someone in a positive light or make her accountable of a particular action, when (2) we want to linguistically communicate our degree (or kind) of commitment to a certain claim to our audience, or when (3) we want to offer the hearer indirect evidence for a claim that we support and we want to reinforce.¹³ Furthermore, a substantial number of empirical studies indicate that human ascriptions are systematically biased and prone to error, which conflicts with the idea that our efficacy in social situations relies on our capacity to identify others' inner mental states. In contrast, *the social cover view*, as we hope to show, can help us to arrive at an understanding of these cases.

According to the social cover view explored here, the three aforementioned uses of ascriptions reveal a common objective: they serve to protect the social status and reputation of an agent either by avoiding responsibility, using a third person as a source of authority when it comes to supporting a particular reason, or providing rationalizations of a particular action. Human interactions are permeated by social norms, cultural rules, shared values or socially scripted patterns of behavior that facilitate and regulate social encounters (McGeer 2007, 2015; Zawidzki 2013). These normative structures, along with mechanisms of reputation management, reciprocation and costly punishment seem to play a central role in cooperative and collective actions.

¹³ We do not rule out the possibility that there are other related functions. Most importantly, we are not negating that we often speculate about the other inner states with our ascriptions. However, we take these functions to be central and more fundamental.

For instance, theories of indirect reciprocity (Alexander 1987; Nowak and Sigmund 2005), competitive altruism (Barclay and Willer 2007; Van Vugt et al. 2007) and costly signaling (Zahavi and Zahavi 1997) all concur on the idea that reputation facilitates cooperation. They suggest that people cooperate to maintain a good reputation in their social environment, where this reputation, in turn, attracts valuable partners and allies, thus positively affecting their future benefits. Such views explain why people's projection of a positive image would be essential in social interactions. In such a context, possessing different strategies to rehabilitate our social status, avoiding possible sanctions or expressing our compliance with norms seems to be an indispensable ability to include in our social repertory. Accordingly, we suggest that the main function of mindreading is not to guide us in discovering the inner states of a subject but to provide different explanations, rationalizations, exculpations, anticipations or justifications for avoiding responsibilities or to cover the social status of the interpreter or the interpreted. Mental state ascriptions are not aimed at the inner mental processes of a subject. Rather, they are ways to attribute conceptual and socio-normative commitments which can cover the social status of a subject.

In this approach, one might expect that people are inclined to protect their own reputation and status over others', which coheres with Malles et al.'s findings of an actor asymmetry in intentional explanation. Further, it explains why people use more ascriptions when asked to put a third person in a positive light. In philosophical literature, other researchers argue a similar point (Andrews 2012, 2015; Fernández Castro 2017; Hutto 2004; Zawidzki 2013). Mental ascriptions serve to rehabilitate social statuses when people deviate from canonical patterns or, to put it in another way, when they carry out counter-normative actions that violate the societal norms which are widely accepted in the community. As Zawidzki (2013, p. 219) claims, our ascriptions serve to restore our social status and avoid potential conflict by making explicit our commitments with rational and social norms.

Although we agree with these authors, we suggest that the protective functions of mental ascriptions are broader than these philosophers have claimed. We often pursue ascriptions with other social purposes that cannot be restricted to those of justification that, nevertheless, have a similar social significance. Consider, for instance, the examples of third-person ascriptions that provide indirect evidence. Providing indirect evidence serves the obvious purpose of giving more credibility to one assertion or claim by exercising the authority of a third person without putting one's own credibility at risk (Gauker 2003; Tooming 2016; Van Cleave and Gauker 2010). Similarly, as the evidence in linguistics indicates (section 5), the parenthetical use of mental verbs conveys a pragmatic meaning that mitigates or modulates our degree of commitment to or uncertainty towards a particular statement. Signaling a low degree of confidence can help us to avoid possible future sanctions if the statement turns out to be false.

Finally, the socio-protective function of avoiding responsibility and sanctions does not require our ascriptions to identify the real causes or internal processes of others. Instead, one could expect people to prioritize the social status and responsibility of the relevant subject over the truth of their inner states. In fact, one might expect explanations and justifications to reflect self-oriented and in-group biases. For instance, the Group-Serving Attributional Bias (Brewer and Brown 1998; Pettigrew 1979), where subjects judge the success of members of other groups as being the result of situational factors out of their control (e.g., luck), while the success of members of their own groups is judged/regarded as being the result of intelligence or talent. Similarly, one

must expect propositional attitude ascriptions to be aimed at protecting ourselves and those who concern us, presenting them and us as well-regulated agents that respect the normative social structures that populate our niche. In this sense, although some of the biases presented in the previous sections (see especially section 4) can be depicted as compatible with the epistemic view, we believe that there are reasons to motivate the idea that they reflect our tendency to project a positive image of ourselves at the expense of its possible epistemic function.

8 Possible Objections

We have offered a way of understanding some empirical studies that are often overlooked in the debate regarding mindreading and mental ascription. Understanding ascriptions as tools for persuasion and social covering facilitates the accommodation of these data. However, our proposal does not hold that persuasion and social covering are the only functions of mental states ascriptions or that we never recruit mental state attributions for epistemic goals. Instead, we claim that using mental state ascriptions for covering is more primary than these other uses. Given that, there are two objections we would like to raise against/mention regarding the social cover view. First, one may object that our proposal is, in principle, incompatible with the developmental findings of so-called implicit false-belief tasks. Second, one could argue that studies into adults' mindreading jeopardize some assumptions of our proposal.

To address the first problem, let us introduce the implicit false-belief task. In 2005, Onishi and Baillargeon (2005) published a set of experiments that seemed to show that 15-month-old infants succeed in a non-verbal version of the false-belief task by measuring looking time. In the experiment, the children were exposed to a change-location scenario.¹⁴ During the experiment, Onishi and Baillargeon measured the looking time of infants to test their reactions. Looking-time measurement is a standard paradigm in developmental psychology. It works under the assumption that babies will look longer at an event when it violates their expectations. In this case, the infants look longer when the person, who was not present when the object was relocated, picks up the object. The infants' sensitivity to false belief was correlated not only with their looking time but also neural responses (Southgate et al. 2007) and helping behavior (Buttelmann et al. 2009; Knudsen and Liszkowski 2012).

Arguably, the results of these experiments can challenge the social cover view. If children ascribe beliefs before they can explicitly use mental verbs for the sake of social covering, then our view is incoherent with the developmental trajectory of mindreading abilities. The most plausible solution to this challenge is to defend a deflationist interpretation of the implicit test results.¹⁵ Such a deflationist

¹⁴ In this task, a child is exposed to a scenario where a character, Maxi, puts chocolate into a cupboard x. When Maxi is not present, his mother moves the chocolate from x into a cupboard y. In the standard task (Wimmer and Perner 1983), children have to indicate the box where Maxi will look for the chocolate when he returns. Only when the child is able to represent Maxi's wrong belief, is he able to point correctly to box x.

¹⁵ Some of the studies involving implicit mindreading have recently failed to replicate (e.g., Kulke et al. 2018). Thus, the debate about how to interpret these tasks may therefore turn out to be futile. However, given the importance of the literature discussing these findings in the last ten years and waiting for more conclusive evidence, we believe the objection is still worth discussing.

interpretation claims that the tasks do not test mental state ascriptions but low-level or sub-mentalizing capacities, like interest contagion (Falck et al. 2014), rational anticipation (Zawidzki 2011) or statistical learning anticipatory mechanisms (Ruffman and Perner 2005). For instance, Ruffman and Perner (2005) contend that children's anticipatory looking behavior may rely on rule-based knowledge concerning what agents ought to do depending on their situations. As such, the social cover view might be compatible with a non-mentalistic understanding of implicit false-belief tasks, and thus with the standard view that situates mindreading later in development (see e.g., Rakoczy 2015). However, unlike the epistemic view, we believe that mindreading capacities are to be associated with social cover capacities. Besides the evidence in section 5, this claim is also supported by the evidence that indicates that children can pass explanatory versions of the false-belief task before they can pass the anticipatory versions (Bartsch and Wellman 1989; Robinson and Mitchell 1995). In these experiments, Bartsch and Wellman presented 3-year-old children, 4-year-old children and adults with different short descriptive theories they had to explain: 'Jane is searching for her kitten under the piano. The kitten is hidden under the chair; why do you think she is searching there?' Then, the different answers were analyzed. They found that 3-year-olds were able to explain such behavior in terms of false belief even when they failed to accomplish predictive versions of the task. These findings do not only support the interpretation that predictive explicit FB-task results are problematic, but it reinforces the idea behind the social cover view. Another source of indirect evidence supporting the social cover view is the findings that children exhibit deceptive behavior early during development (Newton et al. 2000; Reddy 2007). In particular, 2.5-year-old children exhibit *bravado deceptive behavior* "where the child is either denying the experience of the pain being intentionally inflicted on him or her by a parent, or denying the desire for something that he or she has no further hope of obtaining, or re-describing a reality which was detrimental to the child's ego and image of competence" (Reddy 2007, p. 364). To see an example, consider the report of a parent of one episode involving the alteration of the past:

It was early morning and S (sibling) had been staying overnight at a friend's. I was still in bed when R came in and asked me 'Shall we go and pick up S in a minute?' Contradicting him, I said 'No, no. Not in a minute. Later.' R exclaimed 'I said 'shall we go and pick up S later on', that's what I said. I said 'shall we go and pick up S later on'.' This incident struck me and after a few minutes I asked him again 'What did you say?' He answered 'I already said, 'shall we go and pick up S later', I said'. (Newton et al. 2000, Study 2)

Children that exhibit these behaviors are not confused about reality, as manifested by the fact that this example was not isolated or that they acknowledged the lie with a laugh when it failed. This suggests that they were aware of the alteration of reality for deceptive purposes, what indicates an understanding of false beliefs. But also, as the social cover view emphasizes, *bravado* lies are motivated to an ego-defense," a way to protect their own image and competence. Of course, these findings are far from being conclusive. Testing the social cover view requires systematic experiments

demonstrating that scenarios where we need to protect our own or others' social image afford the type of ascriptions and explanations presented above. However, we believe, these findings regarding deceptive behavior and explanatory false-belief tasks, along with the findings presented in section 5, provide some consistent evidence supporting the social cover view.

The second objection is related to different experimental findings regarding adult mindreading. In a recent paper, Evan Westra (2017a) has argued that empirical findings regarding spontaneous mindreading in adults jeopardize what he calls the skeptical views¹⁶ (Bermudez 2003; Morton 1996; Zawidzki 2013) on mindreading, which share substantial claims with the view we are proposing in this article. The rationale for this claim is that such views seem to be committed to the idea that “the attribution of “full-blown propositional attitudes such as beliefs is generally quite slow and effortful, places heavy demands on attention and working memory, and likely requires fairly advanced linguistic abilities” (p. 3). However, Westra claims, recent studies regarding attribution of 2 level perceptual states suggest that mindreading is more spontaneous than these views claim (Elekes et al. 2016; Surtees et al. 2016). For instance, Elekes and her colleagues asked subjects to judge whether or not a number on a screen lying flat in front of them was the same as the number they heard in an audio recording. The subjects could perform the task alone (individual condition) or with another participant performing a similar task (joint conditions). In the joint conditions, the other participant could perform exactly the same tasks (perspective dependent condition) or have to consider the color of the number (perspective independent condition). The crucial point is that the subjects only had to do their own tasks. The experimenters found that the subjects in the perspective dependent condition were slower in the joint condition when performing the perspective-dependent task *and* with the numbers 2, 5, 6 or 9 on the screen, that is, the numbers whose values differed on the basis of perspective. These findings seem to support the idea that mindreading can be spontaneous, since performing the task was not dependent on the other participant's beliefs. Furthermore, Westra argues, this evidence indicates that mindreading is not slow and effortful as the skeptical views suggest.

First, the interpretation of experimental evidence that Westra appeals is controversial. Although it is true that Elekes et al. argue that their experimental evidence suggests that there are instances of “spontaneous perspective taking,” in the original experiments, the 2-Level spontaneous attribution occurred when the participants knew that the partner was performing the same task. Furthermore, in another study conducted by Surtees et al. (2016), they found that the perspective-taking emerged when the subjects know that the partner was performing a task with the same features as their own. Taken together, these findings suggest that perspective taking is sensitive to different contextual factors. However, given that it is controversial which kind of contextual factors drive the attribution, whether the relevance of the task or the partner's attention, (see Elekes et al. 2017, p. 612), we cannot determinate whether the attribution is spontaneous, or the subjects intentionally control the attribution when they consider that it is relevant to the particular task (Surtees et al. 2016). However, even if we accept the

¹⁶ Westra also targets the 2-systems views (Apperly 2011; Bohl and Bos 2012; Butterfill and Apperly 2013; Wellman 2014), according to which the two types of FB-tasks (implicit and explicit) reflect two systems of mindreading. One inflexible, fast implicit system and one slow, explicit verbally mediated system.

interpretation of spontaneity, such evidence is not necessarily incompatible with our theory. As we said before, we do not take the social cover view as proposing that only the positive image projection contexts presented above could afford mental states attributions. Although we take social cover function to be primary, once attributional skills are acquired, we expect subjects to be able to recruit such skills for different purposes. In particular, we could expect that subjects would recruit mindreading spontaneously when others' perspective is relevant for a given task. Indeed, the social cover view must anticipate that certain social factors related to protection management will afford mindreading more spontaneously than others (especially in children). However, in the absence of evidence consistent with this, we can only speculate at this point.

Second, regardless of the controversy related to these experiments, our main point against Westra's objection is, however, the validity of the first premise of his argument. We disagree with Westra on the idea that our view is committed to the idea that mental ascriptions are slow or effortful. Our key point is that epistemic uses of ascriptions are rare in nature and, expectably, harder for the agent than social cover uses. However, this does not mean that such uses cannot be spontaneous or even that they cannot be fast and effortless given the relevant contextual factors. Again, we believe that more experimental evidence is required to clarify this point, but, as Westra himself acknowledges, we believe that the efficiency or effort exhibited in mindreading tasks strongly depend on the ecological validity of the experimental setting. Such an ecological validity would be dependent on different factors, and we expect, especially in the case of adults, such validity not to be dependent only on considerations regarding persuasion, justification or exploiting indirect evidence. However, we expect that such contexts would make the mindreading task especially fast and effortless in comparison with other non-social contexts, especially in the case of children, since we expect them not to be able to properly recruit mindreading for different purposes yet.

9 Conclusions

In this paper, we have argued that there are good empirical reasons to question the epistemic view of mindreading. According to this view, the overall function of mindreading is to gain knowledge of or identify the causal mental states that bring about others' behavior. We have presented different findings in social psychology that demonstrate that humans are systematically biased and prone to error when identifying mental states. Such findings question a fundamental rationale behind the view; namely, that without having accurate and efficient access to other mental states, our capacity for coordination and understanding would remain unexplained. Furthermore, we have presented three different sets of findings that suggest that humans consistently ascribe mental states with the aim of portraying themselves and others positively, indicating a low degree of commitment towards a claim or presenting indirect evidence. These functions, we have claimed, are not only difficult to make coherent with the epistemic view, but they also seem to hold a different perspective on mindreading.

This perspective we have called the social cover view, putting the capacity to ascribe mental states at the service of social protection. Humans ascribe mental states in order to protect their social status, avoiding conflicts or punishment, or persuade others of

different motives. At the end of the paper, we have presented and replied to two possible objections to our view. Although an in-depth analysis is required, we have suggested that this approach has obvious connections with recent proposals claiming that human reasoning is closely connected to persuasion and convincing. Of course, more empirical evidence and conceptual development is necessary, but we believe there are enough similarities between the two projects to explore such an avenue of research. In addition, more empirical studies and conceptual clarification are also required to directly test the social cover view and reframe the available empirical findings within social cognition.

Acknowledgements We are thankful to Agustín Vicente and Andreas Falck for reading previous drafts of this paper and for their valuable comments and suggestions. We are grateful as well for the unconditional philosophical and emotional support to the members of the Granada Gang.

Funding This paper has been funded thanks to the FPI Predoctoral Fellow “BES-2017- 079933”, the Projects FFI2015-65953-P and FFI2016-80088-P funded by the Spanish Ministry of Economy, the project ANR-16-CE33-0017 funded by The French National Research Agency and the FiloLab Group of Excellence funded by the Universidad de Granada.

References

- Aijmer, K. (1997). I think: An English modal particle. In T. Swan & O. Westik (Eds.), *Modality in Germanic language: Historical and comparative perspectives* (pp. 1–48). De Gruyter Mouton.
- Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Alvarez, M. (2010). *Kinds of reasons: An essay in the philosophy of action*. Oxford: Oxford University Press.
- Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology*. Cambridge: MIT Press.
- Andrews, K. (2015). The folk psychological spiral: Explanation, regulation, and language. *The Southern Journal of Philosophy*, 53, 50–67.
- Apperly, I. A. (2011). *Mindreaders: The cognitive basis of “theory of mind”*. New York: Taylor and Francis.
- Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences*, 274, 749–753.
- Baron-Cohen, S. (1995). *Mindblindness an Essay on Autism and “Theory of Mind”*. Cambridge: The MIT Press.
- Bartsch, K., & Wellman, H. (1989). Young children’s attribution of action to beliefs and desires. *Child Development*, 60(4), 946–964.
- Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. Oxford: Oxford University Press.
- Bergamaschi Ganapini, M. (2019). *Confabulating Reasons. Topoi*, online first, 1–13. <https://doi.org/10.1007/s11245-018-09629-y>
- Bermudez, J. L. (2003). The domain of folk psychology. *Royal Institute of Philosophy Supplement*, 53, 25–48.
- Bohl, V., & Bos, W. (2012). Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of type 1 and type 2 processes. *Frontiers in Human Neuroscience*, 6, 274.
- Brewer, M. B., and Brown, R. J. (1998). Intergroup relations. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 554–594). New York, NY, US: McGraw-Hill.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2), 337–342.
- Butterfill, S., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606–637.
- Carruthers, P. (2006). *The architecture of the mind*. New York: Oxford University Press.
- Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. *Synthese*, 192, 1–16.
- Cleave, M. V., & Gauker, C. (2010). Linguistic practice and false-belief tasks. *Mind & Language*, 25(3), 298–328. <https://doi.org/10.1111/j.1468-0017.2010.01391.x>.
- Dudley, R. E. (2017). The role of input in discovering presuppositions triggers: Figuring out what everybody already knew (Doctoral dissertation).

- Dudley, R. E. (2018). Young children's conceptions of knowledge. *Philosophy Compass*, 13(6), 1–18.
- Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry*, 10(1), 1–11.
- Elekes, F., Bródy, G., Halász, E., & Király, I. (2016). Enhanced encoding of the co-actor's target stimuli during a shared non-motor task. *The Quarterly Journal of Experimental Psychology*, 69(12), 2376–2389.
- Elekes, F., Varga, M., & Király, I. (2017). Level-2 perspectives computed quickly and spontaneously: Evidence from eight- to 9.5-year-old children. *British Journal of Developmental Psychology*, 35, 609–622.
- Falck, A., Brinck, I., & Lindgren, M. (2014). Interest contagion in violation-of-expectation-based false-belief tasks. *Frontiers in Psychology*, 5, 23.
- Ferguson, C. K., & Kelly, H. H. (1964). Significant factors in overevaluation of own-group's product. *Journal of Abnormal and Social Psychology*, 69, 223–228.
- Fernández Castro, V. (2017). Regulation, normativity and folk psychology. *Topoi*. <https://doi.org/10.1007/s11245-017-9511-7>.
- Fernández Castro, V. (2019). Justification, conversation and folk psychology. *Theoria*, 34(1), 75–91. <https://doi.org/10.1387/theoria.18022>.
- Fernández Castro, V., & Heras-Escribano, M. (2019). Social Cognition: a normative approach. *Acta Analytica*. <https://doi.org/10.1007/s12136-019-00388-y>.
- Fiebig, A., & Coltheart, M. (2015). Various ways to understand other minds. Towards a pluralistic approach to the explanation of social understanding. *Mind & Language*, 30(3), 235–258.
- Fodor, J. (1992). A theory of the child's theory of mind. *Cognition*, 44(3), 283–296.
- Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind*. London: Routledge.
- Gauker, C. (2003). *Words without meaning*. Cambridge, MA: MIT Press.
- Gilbert, D. T., Krull, D. S., & Pelham, B. W. (1988). Of thoughts unspoken: Social inference and the self-regulation of behavior. *Journal of Personality and Social Psychology*, 55(5), 685–694.
- Goddard, C. (2003). Thinking across languages and cultures: Six dimensions of variation. *Cognitive Linguistic*, 14(2/3), 109–140.
- Goldman, A. I. (1989). Interpretation psychologized. *Mind & Language*, 4, 161–185.
- Goldman, A. I. (2006). *Simulating minds*. New York: Oxford University Press.
- Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.
- Gopnik, G. (1996). Theories and modules; creation myths, developmental realities, and neurath's boat. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (p. 169). Cambridge: University Press.
- Gopnik, G., Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Gordon, R. M. (1996). Radical simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 11–21). Cambridge: Cambridge University Press.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41.
- Hannikainen, I. (2018). Ideology between the lines: Lay inferences about scientists' values and motives. *Social Psychology and Personality Science*, 20(10), 1–10.
- Hansen, M. B., Petersen, E. N., Poulsen, A., & Salès-Wuillemin, E. (2017). When belief ascriptions are about more than what is on someone Else's mind. *Discourse Processes*, 54(8), 655–669. <https://doi.org/10.1080/0163853X.2016.1150042>.
- Hazlett, A. (2010). The myth of Factive verbs. *Philosophy and Phenomenological Research*, 80, 497–522. <https://doi.org/10.1111/j.1933-1592.2010.00338>.
- Heal, J. (1998). Co-cognition and off-line simulation: two ways of understanding the simulation approach. *Mind & Language*, 13, 477–498.
- Hirschfeld, L. A. (2013). The myth of mentalizing and the primacy of folk sociology. In M. R. Banaji & S. A. Gelman (Eds.), *Oxford series in social cognition and social neuroscience. Navigating the social world: What infants, children, and other species can teach us* (pp. 101–106). New York, NY, US: Oxford University Press.
- Hutto, D. (2004). The limits of spectatorial folk psychology. *Mind & Language*, 19, 548–573.
- Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the time course of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, 30(10), 1267–1280.
- Jussim, L. (2012). *Social perception and social reality. Why accuracy dominates bias and self-fulfilling prophecy*. Oxford: Oxford University Press.

- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kalish, C. W. (2012). Generalizing norms and preferences within social categories and individuals. *Developmental Psychology*, 48(4), 1133–1143.
- Kiparsky, P., & Kiparsky, C. (1970). Fact. In M. Bierwisch & K. Heidolph (Eds.), *Progress in linguistics*. The Hague: Mouton.
- Knudsen, B., & Liszowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, 15(1), 113–122.
- Korman, J., & Malle, B. F. (2016). Grasping for traits or reasons? How people grapple with puzzling social behaviors. *Personality and Social Psychology Bulletin*, 42(11), 1451–1465.
- Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science*, 29, 888–900.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lewis, S., Hacquard, V., & Lidz, J. (2017). “Think” pragmatically: Children’s interpretation of belief reports. *Language Learning and Development*, 13(4), 395–417.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, 5(9), 910–916.
- Maibom, H. L. (2007). Social systems. *Philosophical Psychology*, 20(5), 557–578.
- Malle, B. F. (2011). Attribution theories: How people make sense of behavior. In D. Chadee (Ed.), *Theories in social psychology* (pp. 72–95). Malden, MA: Wiley-Blackwell.
- Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93(4), 491–514.
- Mameli, M. (2001). Mindreading, mindshaping, and evolution. *Biology and Philosophy*, 16, 597–628.
- McGeer, V. (2007). The regulative dimension of folk psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 137–156). Dordrecht: Springer.
- McGeer, V. (2015). Mind-making practices: the social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations*, 18(2), 259–281.
- Meltzoff, A. N., Gopnik, A., & Repacholi, B. M. (1999). Toddlers’ understanding of intentions, desires and emotions: Explorations of the dark ages. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 17–41). Mahwah: Lawrence Erlbaum Associates Publishers.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
- Mercier, H., Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Millikan, R. (2004). *Varieties of meaning: The 2002 Jean Nicod lectures*. Cambridge: MIT Press.
- Morton, A. (1996). *Philosophy in practice: An introduction to the Main questions*. Malden, MA: Wiley-Blackwell.
- Moscovitch, M. (1995). Confabulation. In D. L. Schacter (Ed.), *Memory distortion* (pp. 226–251). Cambridge: Harvard University Press.
- Newton, P., Reddy, V., & Bull, R. (2000). Children’s everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology*, 18(2), 297–317.
- Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Oxford University Press.
- Nickerson, R. S. (1988). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175.
- Norman, A. (2016). Why we reason: Intention-alignment and the genesis of human rationality. *Biology and Philosophy*, 31(5), 685–704.
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291–1298.
- Nuyts, J. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*. John Benjamins.
- Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, 46(2), 315–324.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258.
- Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind’. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1480), 731–744. <https://doi.org/10.1098/rstb.2006.2023>.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

- Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, 5(4), 461–476.
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Science*, 11(1), 37–43.
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28(3), 369–381.
- Rakoczy, H. (2015). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese*, 194, 689.
- Reddy, V. (2007). Getting back to the rough ground: deception and 'social living'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 621–637.
- Robinson, E. J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development*, 66(4), 1022–1039.
- Rubio-Fernandez, P., & Geurts, B. (2012). How to pass the false-belief task before your fourth birthday. *Psychological Science*, 24(1), 27–33.
- Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Science*, 9(10), 462–463.
- Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, 34(3), 265–293.
- Rule, N. O., Ambady, N., & Adams, R. B., Jr. (2009). Personality in perspective: Judgmental consistency across orientations of the face. *Perception*, 38, 1688–1699.
- Sabbagh, M. A., & Callanan, M. A. (1998). Metarepresentation in action: 3-, 4-, and 5-year-olds' developing theories of mind in parent-child conversations. *Developmental Psychology*, 34, 491–502.
- Sanchez-Curry, D. (2018). Beliefs as inner causes: The (lack of) evidence. *Philosophical Psychology*, 31(6), 850–877.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Scheibman, J. (2001). Local patterns of subjectivity in person and verb type in American English conversation. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 62–89). Amsterdam: John Benjamins.
- Schneider, S., & And Glikman, J. (2015). Origin and development of French parenthetical verbs. In S. Schneider, J. Glikman, & M. Avanzi (Eds.), *Parenthetical verbs* (pp. 163–188). Berlin: de Gruyter.
- Shatz, M., Wellman, H., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, 14, 301–321.
- Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, 117, 1034–1056.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587–592.
- Spaulding, S. (2016). Mind misreading. *Philosophical Issues*, 26(1), 422–440.
- Spaulding, S. (2018). *How we understand others: Philosophy and social cognition*. London and New York: Routledge Focus.
- Stone, T., & Davies, M. (1996). The mental simulation debate: A progress report. In P. Carruthers and P.K. Smith (eds.) *Theories of theories of mind* (pp 119–137). Cambridge, UK: Cambridge University Press
- Strijbos, D. W., & De Bruin, L. C. (2011). Making folk psychology explicit: The relevance of Robert Brandom's philosophy for the debate on social cognition. *Philosophia*, 40, 139–163.
- Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, 150, 43–52.
- Tanney, J. (2013). *Rules, reason and self-knowledge*. London: Harvard University Press.
- Thompson, S. A., & Mulac, A. (1991). A quantitative perspective on the grammaticalization of epistemic parentheticals in English. *Approaches to Grammaticalization*, 2, 313–329.
- Tooming, U. (2016). Beliefs and desires: From attribution to evaluation. *Philosophia*, 45(1), 359–369.
- Urmson, J. O. (1952). Parenthetical verbs. *Mind*, 6(244), 480–496.
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition*, 116(1), 87–100.
- Van Vugt, M., Roberts, G., & Hardy, C. (2007). Competitive altruism: Development of reputation-based cooperation in groups. In R. Dunbar & L. Barrett (Eds.), *Handbook of Evolutionary Psychology* (pp. 531–540). Oxford: Oxford University Press.
- Wellman, H. R. (2014). *Making minds: How theory of mind develops*. NY: Oxford University Press.
- Westra, E. (2017a). Spontaneous mindreading: A problem for the two-systems account. *Synthese*, 194(11), 4559–4581.
- Westra, E. (2017b). Stereotypes, theory of mind, and the action-prediction hierarchy. *Synthese*. <https://doi.org/10.1007/s11229-017-1575-9>.
- Wierzbicka, A. (2006). *English: Meaning and culture*. New York: Oxford University Press.

- Wilson, G. and Shpall, S. (2016). Action. The Stanford Encyclopedia of *Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2016/entries/action/>>. Accessed 20 April 2019
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81(5), 815–827.
- Zahavi, D. (2011). Empathy and direct social perception: A phenomenological proposal. *Review of Philosophy and Psychology*, 2(3), 541–558.
- Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. Oxford: Oxford University Press.
- Zawadzki, T. W. (2011). How to interpret infant socio-cognitive competence. *Review of Philosophy and Psychology*, 2(3), 483–497.
- Zawadzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge: MIT Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.