



Shaping your own mind: the self-mindshaping view on metacognition

Víctor Fernández-Castro¹  · Fernando Martínez-Manrique² 

Published online: 17 January 2020
© Springer Nature B.V. 2020

Abstract

Starting from Proust's distinction between the self-attributive and self-evaluative views on metacognition, this paper presents a third view: self-mindshaping. Based on the notion of mindshaping as the core of social cognition, the self-mindshaping view contends that mindshaping abilities can be turned on one's own mind. Against the self-attributive view, metacognition is not a matter of accessing representations to metarepresent them but of giving shape to those representations themselves. Against the self-evaluative view, metacognition is not blind to content but relies heavily on it. We characterize our view in terms of four issues that, according to Proust, distinguish the previous approaches, namely, whether metacognitive mechanisms are the same as those employed to access other minds, whether metacognitive control requires conceptual representation, whether metacognition is propositional, and whether metacognitive access is linked to mental action. After describing some of the mechanisms for self-mindshaping, we show how this view regards metacognition as (1) grounded on social interaction mechanisms, (2) conceptually driven, (3) possibly, but not necessarily, propositional, and (4) engaged in the practical regulation of mental states. Finally, we examine the prospects for the primacy of self-mindshaping as the primary metacognitive function. We argue that self-attributive processes typically subserve the practical goals emphasized by the mindshaping view, and that the evaluative role played by procedural metacognition can be grounded on social cues rather than on experiential feelings. Even if this is not enough to claim the primacy of self-mindshaping, it still appears as a third kind of metacognition, not reducible to the other two.

Keywords Metacognition · Self-mindshaping · Self-attribution · Self-evaluation · Metarepresentation

✉ Fernando Martínez-Manrique
fmmanriq@ugr.es

Víctor Fernández-Castro
vfernandezcastro@gmail.com

Extended author information available on the last page of the article

1 Introduction

A key aspect of human agency and self-governance is metacognition, which can be characterized, in general terms, as the capacity for directing one's cognitive processes towards one's own cognition. Although it is difficult to make this characterization more precise without biasing it towards a specific view, it seems that there are two extremes that should be avoided. On the one hand, the idea that metacognition has to do with any cognitive process that intervenes in another cognitive process, which would mean that too many mental mechanisms would count as metacognitive. On the other hand, the idea that metacognition has to do with some sort of high-level knowledge of our mental processes and states –e.g., knowledge about how I remember, how I reason, or how I perceive– which would mean that perhaps only a few privileged individuals could be actually capable of metacognize. In between these poles there is room for regarding as metacognitive a set of abilities devoted to accessing cognitive states and processes with the purpose of monitoring and regulating them, so as to make them more effective. There seem to be metacognitive abilities of this kind involved in apparently very different cognitive processes, such as memory, reasoning, or deliberation, so it is not easy to characterize what those abilities have in common. Proust (2013) proposed that there are two basic ways to offer a general characterization. One of them regards metarepresentational abilities as the crucial component: The way in which humans flexibly modify and regulate their cognition is based on the ability to represent their own thoughts, i.e., on metarepresentation. Proust labels this position as the *self-attributive view*, and calls into question its main tenet. Metacognition, she argues, does not require the manipulation of metarepresentations. On the contrary, metacognitive capacities can be carried out by procedural mechanisms that are blind to the content of mental states and that simply evaluate whether the states and processes involved are doing what they are expected to do. This is what she calls the *self-evaluative view*.

The aim of this paper is to present a third alternative characterization of metacognition, which we will call the *self-mindshaping view*. Its basic claim is that there are forms of metacognition that are based on the same kinds of mechanisms we use to shape others' minds. In some respects, this approach can be regarded as a middle way between the attributive and evaluative views: it shares with the former an emphasis on content-driven processes, and with the latter an emphasis on evaluation. Yet, we contend that it is not a variety of either of them. In contrast with typical instances of the self-attributive view, self-mindshaping holds that metacognition is not primarily a matter of accessing representations that are “already there” to be re-represented: instead, in many relevant cases it is a matter of giving shape to those representations themselves. In contrast with the self-evaluative view, metacognition is linked to social cognition and conceptual abilities.

The paper will be structured as follows: First, we present the distinction between self-attributive and self-evaluative views of metacognition in terms of four questions posed by Proust to highlight their points of disagreement. Our aim will be to show that, against Proust, their answers do not cluster in two distinct camps. It is possible to answer some questions aligning with one camp and others with the opposite camp, and what matters is the way in which the question is answered. Then, we will present the notion of mindshaping, a recent approach to social cognition that attempts to provide an alternative to the widespread idea of mindreading. Our aim will be to describe briefly

some mechanisms that characterize mindshaping, as the background for the self-mindshaping view we want to sustain. Afterwards we show how those mechanisms can be put into service for metacognitive purposes, and we explain how the self-mindshaping view answers the four questions posed previously, so as to highlight its differences with self-attribution and self-evaluation. Finally, in the last section we address the issue of the relation between self-mindshaping and the other two alternative accounts. Borrowing a term from Peters (2019), we examine the prospects of the *primacy* of self-mindshaping as the primary metacognitive function. On the one hand, we argue that self-attributive processes typically subserve the practical goals emphasized by the mindshaping view. On the other hand, we argue that the evaluative role played by procedural metacognition can be grounded on social cues, rather than on experiential feelings. The upshot is that even if there are mechanisms such as those proposed by self-attributive and self-evaluative views, they make better sense in the context of the normative conceptual regulation of one's own mind that we call self-mindshaping. Even if this is not enough to claim the primacy of self-mindshaping, we conclude that at the very least it appears as a third kind of metacognition, not reducible to the other two.

2 Two approaches to metacognition

Metacognition is roughly a matter of directing cognition toward itself. In an attempt to capture in a single term the core idea behind this notion we will resort to the term 'access'.¹ Metacognition is the ability to access our own mental states or processes, where 'access' can be understood in different senses. One way to access a mental state is to know it, i.e., to be able to tell what kind of state it is, or what its content is. A second sense of 'accessing' a mental state, however, is to have the capacity to control it, much as when one is capable to access a lever by reaching one's hand. These two notions correspond roughly to the two main approaches to metacognition distinguished by Proust (2010, 2013): the self-attributive and self-evaluative views.

The self-attributive view (Perner 1991; Leslie 1994; Goldman 2006; Carruthers 2009, 2011) is essentially a representationalist view of our access to mental states, according to which metacognition requires forming representations about a mental state in order to perform other actions such as controlling or regulating the state in question. In contrast, in the self-evaluative view (Proust 2013) it is possible to access a mental state without representing the state at all, in order to control or regulate it in some respect. In particular, the self-evaluative view envisages a particular variety of control – namely, evaluation – as a basic metacognitive ability that is independent from any putative capacities of knowing about our mental states.

Proust (2013, chapters 2 & 3) presents two sets of four claims that characterize each view. We think that it can be useful to express those claims as answers to four different questions, which can be then used as a heuristics to position other possible approaches,

¹ As an anonymous reviewer observes, there are alternative terms, such as 'use', which in certain cases could avoid the connotations that 'access' involves. We agree that it is difficult to find a single term that covers all cases and we hope that our explanation of how to understand it suffices the reader to avoid the undesired connotations.

like the self-mindshaping view that we will motivate. To this end, we are going to phrase those questions in a slightly different way from hers²—resorting to the notion of ‘accessing’ a mental state that we just presented— but keeping intact the gist of her distinction.

- (1) *Are the mechanisms for accessing our mental states the same as those for accessing the mental states of others?* The self-attributive view deems that there is a strong connection between metacognition and social cognition. To know about someone’s mental states we have mindreading abilities that allow us to represent her mental contents and her attitudes to them, so as to conclude, e.g., that she believes (desires, hopes...) that p . Knowing about one’s own mental state would involve basically the same mechanism, turned toward one’s own mental representations. On the contrary, the self-evaluative view regards metacognition as a functionally distinct system that is not dependent on mindreading mechanisms. It is a system devoted to the evaluation of cognitive processes, and this evaluative activity is different—more direct, to begin with—from any means I may have of evaluating the mental states of others. This is what Proust calls the *exclusivist* view of metacognition. In fact, as there are many different cognitive abilities (decision-making, motor abilities, memory), one may conjecture that each of them could have its own evaluative mechanism associated to it and fully independent of mindreading.
- (2) *Does access as control require access as representation?* The self-attributive view holds that any ability to control our mental states—including the ability to evaluate them— demands a correct representation of the mental state in question. Although it is possible to know a mental state without being able to control it, it is not possible to control a state that one does not really know. Metacognition is thus content-guided. The self-evaluative view rejects this answer: there are at least certain kinds of control processes that do not involve forming a representation of the state to be controlled. Consider, for instance, control of a memory search: it involves the ability to know that one is capable of remembering a certain fact, so as to actually perform the search, but it does not involve access to the fact itself. What Proust is proposing is a basic procedural architecture composed of mechanisms that evaluate the processes in question and send re-afferent signals (what Proust calls noetic feelings) to inform the rest of the system whether or not the cognitive process in question is working as expected. These mechanisms perform metacognitive functions in which there is no need to access the contents of the mental states involved, only the competence of the processes in charge.
- (3) *Does access to our mental states rely on propositions?* As the self-attributive view regards metacognition in terms of access to represented mental states, and as the typical mental states to be represented are propositional attitudes, mechanisms that access those states will be metarepresentational mechanisms capable of representing the propositional content of the attitude. So, if I desire that p , accessing this state involves forming the metarepresentation that I have the desire that p , and this representation has p as its component. The self-evaluative view, however, rejects a propositional view of metacognition. Instead, for Proust the

² For explanatory purposes, we will also reverse the order in which she presents claims 2 and 3.

foundation of metacognition lies in what she calls “noetic feelings”. These are defined as:

Generated while agents are trying to perform a cognitive task. In the cases of tasks to be performed, feelings are generated when the agents do not immediately retrieve or perceive an element relevant to the task (e.g. ‘What is the name of X?’). Such feelings predict potential success or error in the current task (prospective noetic feelings, e.g. feelings of knowing). In the cases when a cognitive task has just been performed, the feelings indicate whether the output matches internal standards of correction (retrospective feelings, e.g. feeling of being right). (Proust 2013, p. 318).

In sum, noetic feelings indicate whether the output of the first-order mental action is accurate or not and trigger the appropriate response. They are neither representational nor propositional, even if they work on the outputs of representational mechanisms. For instance, Proust (2013, pp. 121–122) discusses a possible representational mechanism (feature-based representational systems) that produces nonpropositional representations of the form ‘Poor (excellent, etc.) A-ing affordance!’ where the noetic feeling (poor) evaluates or commands whether or not the agent must proceed with the cognitive action (a-ing = remembering, comparing, categorizing...) given the information she has on the task.

To be sure, Proust does not reject a role for metarepresentation in metacognition but she limits it to high-level forms of control, which constitute what she calls *analytic metacognition*. This is still an evaluative process yet it has different epistemic norms that provide explicit and flexible ways of controlling one’s own cognitive processes. Procedural and analytic metacognition coexist in adult humans, in the sense hypothesized by dual-system theory (Evans and Frankish 2009). The former would be based on system-1 mechanisms and remains active throughout life even though it can be partially overridden by the latter, analytic system-2 (Proust 2015). Nevertheless, Proust’s main claims aim not only to guarantee the independence of the former but, as we will see in section 5.2, to argue for a certain primacy of this basic system in metacognition.

- (4) *Is access to our mental states inherently associated to mental action?* Proust (2001, p. 107) regards mental acts as intentional actions that have mental-representational properties as their goals. The goal could be a mental event, state, or disposition, as far as it belongs to the subject. In this sense, remembering or inferring would be typical kinds of mental actions. However, many mental activities can count as mental actions when they have an intentional orientation. For instance, even if perceiving is often an unintentional process, it can be performed in an intentional manner, as in the case of perceiving attentively. Mental actions are thus kinds of doings that are carried out in a controlled way. As any other action, they can result in success or failure. According to the self-evaluative view, metacognition is a matter of assessing how successful I was in doing what I was mentally trying to do, so accessing a mental state always takes place in the context of a particular mental action, and there is a procedural self-evaluation linked to each instance of mental agency. The self-attributive view, in contrast, holds that what we get when we access our (or others’) mental states is a

kind of knowledge, and the knowledge we gain and the action we take are different things. As a consequence, the self-attributive view regards as metacognitive a different range of mental states: Knowing what I am currently perceiving or what emotion I am experiencing would count as instances of metacognition. In contrast, for the self-evaluative view only actions associated with those states would fall under the scope of metacognition, e.g., to evaluate the quality of my perception or the appropriateness of my emotion.

Those are then the questions that broadly differentiate the two camps that Proust distinguishes. There are two observations that we want to make with respect to this analysis. One is that those four questions can be regarded as relatively independent from each other. For instance, one could perhaps accept the claim that the mechanisms for metacognition are basically the same as the mechanisms to deal with other-cognition but reject the idea that those mechanisms are propositionally based or that self-control always demands self-representation. If this is so, then there is room for a number of positions that are intermediate between the self-attributive and the self-evaluative views. It is not clear to us whether any such view would qualify as self-attributive or self-evaluative. The second observation is that the self-evaluative and the self-attributive views differ in how they conceive the notion of *accessing* a mental state, i.e., as accessing its content or as accessing its success. This, in turn, makes them differ in the type of architecture and representational formats they postulate to account for metacognition.

The alternative view of metacognition, self-mindshaping, that we are about to outline capitalizes on these observations. First, we will see that it does not side with either the self-attributive or the self-evaluative view but has a mixed pattern of answers. In particular, like the self-attributive view, it answers affirmatively to question 1, as it regards metacognition as linked to social cognition, and, like the self-evaluative view, it answers affirmatively to 4, as it regards metacognition as linked to mental actions with practical goals. With respect to questions 2 and 3, its emphasis is not on propositional but on *conceptual* metacognition, i.e., metacognitive mechanisms are typically concept-driven and they can be propositional, though not necessarily. Second, and most important, self-mindshaping relies on a notion of accessing a mental state that is not reducible either to epistemic content access or to content-blind evaluative access: it is the idea of accessing as giving shape to a mental state to which you subsequently commit. So the fact that its answers are aligned with either of those positions does not make it a variant of them. Having a third notion of ‘access’ helps to dispel the idea that, even if questions 1–4 are formulated in a yes/no manner, we must choose only between two theories behind the answers. Given this, our point is *not* that self-mindshaping is a hybrid of the attributive and evaluative views but that it is an alternative to them, based on distinct mechanisms. Let us first present the idea of mindshaping and, subsequently, how it can be applied to develop an alternative view of metacognition.

3 Social cognition as mindshaping

The approach to metacognition we endorse in this paper is related to a specific manner of understanding social cognition. In this sense, the view we canvass shares with the

self-attributive view the idea that there is an intimate connection between our metacognitive and social cognitive skills. However, while the self-attributive view starts with the idea that social cognition relies on mindreading capacities that require (meta)representing others' mental states, the self-mindshaping view starts with a radically different understanding of social cognition: the *mindshaping view* (Mameli 2001; McGeer 2007a; Zawidzki 2013). Let us review the main theses behind this approach so as to see how it can be applied to a theory of metacognition.

The standard mindreading approach to social cognition conceives socio-cognitive practices primarily as constituting an *epistemic* enterprise where social creatures attempt to get knowledge of others' mental states for the sake of prediction and explanation (Almagro Holgado & Fernández Castro 2019; Zawidzki 2013, pp. xii-xiii). If the interpreter represents correctly the relevant mental states, then she will be able to provide a more accurate prediction of the interpretee's future behavior, and a more coherent explanation of his past behavior (see, for instance, Fodor 1992; Gopnik and Meltzoff 1997, pp. 13–42; Penn and Povinelli 2007, p. 394). In contrast, the mindshaping view contends that social interactions are primarily *practical*, i.e., normatively mediated actions where social agents deploy different strategies to form and stabilize a myriad of social norms and routines that facilitate coordination. Hence, the main job of the agent is not to represent the others' mental states but to regulate them. As McGeer (2015) puts it:

The regulative view rejects the standard idea that folk-psychology involves a primary capacity for discovering or detecting (pre-existing) mental states; rather it argues that folk-psychology involves a primary capacity for forming and regulating our mental states in accordance with a rich array of socially shared and socially maintained sense-making norms. (McGeer 2015, p. 260).

Acquiring those norms and patterns of behavior facilitates our social interactions by making behavior more transparent for others. Thus, social agents can predict each other because they fulfill normative expectations by regulating their behavior through those expectations. We do not necessarily understand the behavior of others by positing mental states that bring out those courses of behavior. Instead, we assume that others behave according to normalized patterns of behavior regulated by social norms. Inasmuch as those norms are shared they lead to expected practices that make our behavior intelligible to each other.

This regulative function extends to language. Zawidzki contends that “the point of language is to regulate behavior, rather than to offer a window into the independently constituted mental states of interlocutors” (2013, p. 138). This means that even if I can describe you in terms analogous to ‘she believes that *p*’ this is not a representation that was formed after tracking your mental states but a representation of the expectation that the courses of action to which you are committed given the current situation are those consistent with *p*. Similarly, as we will see, in describing myself as ‘I believe that *p*’ I am not describing a mental state that I somehow found in my mind, either in a quasi-perceptual or in an inferential way, but making up my own mind by committing myself to courses of action consistent with *p*. To be sure, to make such descriptions one must resort to linguistic abilities that include the capacity to construct recursive representations and to handle mental state verbs with propositional clauses. Yet it is not necessary to attribute sophisticated mindreading as the basis for such constructions. Rather,

Zawidzki proposes that we should conceive the communicators' capacities as behavior reading guided by certain normative assumptions (2013, p. 146).³

The mindshaping view thus introduces a different perspective on social cognition. Folk psychological abilities encompass mindshaping practices and mechanisms, so it is "a group accomplishment, involving simultaneously interpretive and regulative frameworks that function to shape minds, which these frameworks can then be used to easily and usefully interpret" (Zawidzki 2013, p. xiii). The primary function of social cognition is not to know others' mental states but to mold others' minds and actions such that they become well-behaved agents and, as an outcome, well-predicted and explained too (McGeer 2007a). This involves a switch of focus from mindreading mechanisms to mindshaping mechanisms and practices that operate to sculpt individuals' minds. Some of the mechanisms proposed by Zawidzki (2013) include imitation, natural pedagogy, norm recognition and enforcement, stereotype assumptions, or narratives. They provide ways to enhance the control of homogenization of populations in a way that affects all the interacting individuals, so that they can get involved in complex social projects, predictions and explanations. In the next section we will see how this competence to make sense of others is intertwined with our competence to make sense of ourselves. As we are going to contend, neither of them is a mere epistemic enterprise: both are practical capacities of regulating behavior and cognizing.

4 Self-mindshaping as an account of metacognition

Mindshaping holds the promise of offering a distinctive treatment of metacognition, i.e., one that is not reducible either to the self-attributive or the self-evaluative view presented above. If our minds have resources to shape cognition and our own minds are amenable to being shaped, then it is natural that shaping resources can be turned on one's own mind. So, some of the folk psychological competences we apply to regulate others' behaviors and make it more predictable can be used to control our own behavior and minds. In this sense, the evaluations and expectations that drive our metacognitive processes derive from the normative structures that regulate our social interactions.

Shaping provides a third notion of 'accessing one's mind' that, we contend, is not reducible either to 'access as knowing' or 'access as controlling'. Shaping is both a regulative and formative notion. Shaping is a regulative process inasmuch as it is a way of controlling a particular cognitive trajectory by setting a number of constraints on it. At the same time, shaping helps to give form to the very mental states that will figure in such a trajectory. It is especially regarding this aspect, we believe, that mindshaping can

³ Zawidzki is concerned with providing an alternative story to tell about how the capacity for structurally complex language came into being if it was not to serve the expression of structurally complex thoughts. His own proposal is that such a capacity "is a product of selection, in prehistory, for costly, honest signals of commitment to groups and cooperative endeavors in ritualistic displays" (2013, p. 167). We do not wish to endorse this proposal, or other alternative stories. Our concern is limited to a putative role of language in mindshaping that does not require metarepresenting mental states. Some authors put the focus on language as a capacity to construct external symbols that can be manipulated in a second order cognitive dynamics (Clark 1998). Vierkant (2012) capitalizes on this observation to point out that language allows attention to the content of the linguistic vehicles without understanding the relationship between content and mental states. So mindshaping would require language but not metapresentation. (We will come back to his views on section 5.1).

provide an original approach to metacognition, which appears as a form of self-directed agency based on the capacity to regulate our own minds. On this self-mindshaping view, and contrary to the notion of knowing handled by self-attributive view, accessing our own minds is not primarily based on metarepresenting our mental states. Regulating our mental states only demands generating expectations about what those states ought to be and then living up to those expectations. As McGeer (2007b) argues, “when we are asked to report what’s in or on our minds, we do not track or map internal mental states” (McGeer 2007b, p. 87). Rather, we register the state of the world and attribute ourselves a mental state we ought to display. This attribution employs a linguistic vehicle but this is not an expression of a mental state that we happen to find, it is an expression of a content by which our behavior should abide. So, accessing your mind is not a matter of bringing to light covert pre-existent mental states, but of being actively committed to deploy mental content, perform mental actions, and regulate overt behavior according to certain normative standards.

Despite the emphasis on regulation, shaping is not reducible to ‘access as control’ either, because the kind of regulation it advocates is typically mediated by content. So the mechanisms on which it is based can be distinguished from the content-blind procedural architecture that the self-evaluative view proposes. Some of the mindshaping mechanisms that are recruited for metacognition rely heavily on content—even, as in the case of narratives, on propositional content—as part of the expectation-forming processes that drive the regulation and formation of our cognitive states.

In a nutshell, our basic claim is that the notion of shaping provides a notion of access that can support an approach to metacognition that is an alternative to self-attribution or self-evaluation. Our aim in this section is to characterize the basic elements of this notion of *self-mindshaping* by examining how it fares with respect to the four questions (1–4) that we summarized above. Metacognition will thus appear as: (1) grounded on social interaction mechanisms, (2) conceptually driven, (3) possibly, but not necessarily, propositional, and (4) engaged in the practical regulation of mental states.

4.1 The social roots of metacognition

Like the self-attributive view, and unlike Proust’s self-evaluative view, self-mindshaping rests on the idea that our metacognitive capacities rely on socio-cognitive competences, that is, the way we access our own minds does not differ drastically from the way we access other minds as we interact with other agents. However, as we saw in section 3, the idea of mindshaping offers a way to understand those socio-cognitive capacities that differs radically from the mindreading perspective that is the backbone of the self-attributive view. Now, how does mindshaping turn into self-mindshaping? We will sketch an answer to this question by examining how mindshaping mechanisms work. Actually, it is possible to distinguish two ways in which mindshaping can become self-mindshaping.

The first way is apparent in the structure of mindshaping processes themselves. An important thing to note is that, unlike mindreading, mindshaping is an interactive affair. If I read your mind, I am trivially affected by having a new belief about your mental states, and possibly by the content of the attribution I just made, e.g., that you have racist beliefs. But this is not an interactive process, as it involves unidirectional changes. In contrast, mindshaping appears as a highly interactive process that

inevitably affects both the mindshaper's and the mindshaped's minds: if shaping other minds has to do with establishing and negotiating certain norms that make them more predictable, it is clear that this normative structure affects all the individuals involved. In other words, as mindshaping is a social interactive affair, it becomes as much a way of shaping the other minds as one's own mind, which renders itself more predictable as well. A mechanism in which this two-way effect is particularly clear is imitation. Imitating others is a learning mechanism for oneself but it also plays an important function in coordination of group actions (Tsai et al. 2011). Although imitation is not unique to human beings, what is characteristic of human offspring is that they compulsively "overimitate" adult behavior with a high degree of fidelity, e.g., children imitate causally irrelevant aspects of a goal-directed task (Horner and Whiten 2005; McGuigan et al. 2011). The importance of overimitation from the mindshaping viewpoint is that it facilitates the acquisition of patterns of behavior, which in turn facilitates social interactions. The capacity to regulate behavior according to external models enhances prediction because once we copy some courses of behavior in specific contexts, others will expect us to trigger those behaviors in those contexts.

The second way in which mindshaping can turn into self-mindshaping is related to the fact that mindshaping mechanisms can be intentionally turned towards oneself, becoming effective metacognitive mindshaping tools. Imitation provides again an example when it takes the form of *self-imitation*. Rochat (2002) argues that self-imitation has an *ego* function that provides a primary source of knowledge about oneself and a way to develop self-reflective abilities. However, notice that self-imitation has to do with "the systematic attempt to *reproduce* and *match* previous patterns of self-generated action" (Rochat 2002, p. 88). So, in our view, its role in self-knowledge should not be understood as the basis of a capacity to attribute states to oneself but as a capacity to form those states and to commit oneself to their consequences, just as the mindshaping view claims.

Self-mindshaping can thus occur in an indirect way⁴—as a result of shaping one's interactions with others—and in a more direct, intentional way—aiming a mindshaping mechanism at one's own cognitive processes. These two ways can be observed in other proposed mindshaping mechanisms. Consider, for instance, Vygotsky's model of socially based development of cognitive processes. Working within the *zone of proximal development*,⁵ adults may elicit appropriate children's responses to solve a particular problem—e.g., a jigsaw—by introducing different guides, questions, commands or instructions. This pedagogical mechanism helps to homogenize behaviors in order to make them more predictable, regulating behavior and cognitive processes according to

⁴ Answering a concern from an anonymous reviewer, we want to clarify that by "indirect way" we do not mean "unintentional way". It does not include all actions that facilitate or influence cognition. So turning my head towards a certain stimulus because it became salient should not count as mindshaping; but turning my head towards a stimulus because you made a gesture pointing at it may count as a form of mindshaping inasmuch as you want to make the stimulus salient to me so that both of us perceive it in the same manner and coordinate our attitudes toward it. Of course, intentional processes can exploit automatic mechanisms, e.g., tendency to imitation or conformity to norms. But just as one should not confuse the process of reasoning with inference-drawing mechanisms—that can be automatic—, one should not confuse the process of mindshaping with the mechanisms employed in the process.

⁵ This is "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky 1978, p. 86).

certain external standards in a social niche. Then, the children can reproduce the same instructions in order to solve the problem on their own, internalizing the appropriate structures that facilitate the ways to solve the cognitive task. So the same mechanisms that work in the zone of proximal development can be directed towards one's own mental processes.

The mindshaping view also emphasizes the regulative role of narratives (Hutto 2004; Nelson 2007; Sterelny 2012). Story-telling allows children to grasp patterns from which they evaluate what is permissible or not. This facilitates the learning of shared values, norms and routines that constitute the ideology of their social niche. In other words, narratives function as models of appropriate and normalized behavior for children. As Hutto puts it, narratives “shape our expectations by making us familiar with a vast stock and wide range of ‘ordinary’ situations and the sorts of actions normally related to them” (Hutto 2004, p. 559). In this sense, narratives are a typical means for children to learn the normalized patterns of action that help them to make sense of others. At the same time, as children start to develop inner speech capacities, they become capable of directing their narratives toward themselves, employing them in a number of self-regulative functions (Alderson-Day and Fernyhough 2015; Fernández Castro 2017, 2019; Martínez-Manrique & Vicente 2015).

The mechanisms just presented are but a sample of mindshaping processes and it is not our intention to discuss them at length. What they suggest is that, in spite of being constitutively linked to socio-cognitive capacities, self-directed cognitive processes do not require mindreading. On the other hand, some of those mechanisms demand representational resources, even propositional ones. So we turn to consider the issue of how these resources are managed in the self-mindshaping view.

4.2 The conceptual character of metacognition

The second question (“Does access as control require access as representation?”) is concerned with the representational character of the mental states that our metacognitive capacities monitor. According to the self-attributive view, controlling our mental states requires the agent to represent the mental states in question. In contrast, the self-evaluative view holds that our capacity to evaluate our mental processes can be guided by dynamic cues generated by some feedback mechanisms that control psychological competence on line. Accessing the relevant represented content does not play a role in that control. Here the question of putative metacognition in nonhuman animals plays an important role in Proust's argument. In her view, there are cases in which it is reasonable to attribute them metacognitive abilities, namely, cases in which the animal has to make an assessment of its certainty of being able to carry out a given task. Yet instead of explaining them by means of concept-involving mindreading abilities it is possible to do so in terms of nonconceptual representations. They are procedural representations that provide the creature with an assessment of its cognitive processes.

Now, we agree with Proust that mindreading resources, with their typical reliance on ascriptions of propositional attitudes, are beyond the reach of nonhuman animals. Yet we dispute the idea that it is possible to make sense of many of their metacognitive abilities in purely nonconceptual terms. In social species, there are coordination abilities that demand conceptual metacognitive access of one's position and role in relation to

one's conspecifics. Let us examine this point with an example from hunting strategies in chimpanzees.

Boesch (2002) shows that Tai Chimpanzees can play four different roles in hunting contexts, namely, chaser, ambusher, driver, and captor. The role they play in each case depends on variables such as location or anticipation of the prey's behavior. We contend that such sophisticated behavior requires some metacognitive abilities, and these are better regarded as an instance of self-mindshaping. First, hunting strategies are cooperative tasks that cannot be afforded without a high degree of social expertise. Learning to be a chaser or a driver implies learning a set of patterns of behavior and social norms, which help to regulate the subject's behavior in different contexts. Second, the other hunting partners can use those sets of norms and patterns in order to predict the behavioral responses of each other and behave appropriately: If chimpanzee A expects chimpanzee B to assume the role of driver according to B's location, A can assume the role of basher and thus trigger specific cognitive and behavioral responses. This social interaction is not a matter of mindreading capacities –indeed, it is questionable that chimpanzees possess them (Andrews 2009). Third, assuming a specific role is a form of shaping its own mind, i.e., it is a metacognitive capacity in the sense that it allows the chimpanzee to control its own behavior and cognitive responses according to the appropriate norms and patterns. Even though this ability does not require making complex mental state attributions, it is a concept-involving ability in which the chimpanzee assigns itself and its conspecifics the right labels. So instead of entertaining a complex thought like 'I believe I am a captor', to assume the captor role it is enough to have the concept CAPTOR playing an active role in initiating a pattern of appropriate behavioral responses that, in turn, involve further concepts.⁶ For instance, assuming the captor role generates a set of cognitive responses in line with the belief-like state 'prey at location L', which the subject needs to keep in mind to produce the appropriate behavioral responses. So a plausible explanation of the behavior seems to involve, first, the ability of making generalizations involving concepts like LOCATION or PREY but not the ability to entertain complex beliefs, and second, the metacognitive ability of recognizing itself as playing a certain role, so as to coordinate its behavior with the others. Hence the Tai Chimpanzees case contrasts both with the self-attributive view –it does not resort to metarepresentational states– and with the self-evaluative view –it demands conceptual states.

We contend that the ability to modify certain cognitive strategies according to the normative expectations that each role generates is a metacognitive ability. To see how, notice that meta-cognitive regulation requires executive skills related to monitoring and self-regulation of one's own cognitive activities (Schneider 2008). As we said, a typical example of metacognitive self-regulation is the classical paradigm of measuring the monitorization of certainty and uncertainty in non-human animals (see Couchman et al.

⁶ The notion of concept that is at issue is not as the constituent of a propositional thought but as a content-involving structure that is relatively independent from stimulus. This is a minimal sense of concept that a number of authors posit to account for complex behaviors and instrumental reasoning in certain nonhuman animals (Newen and Bartels 2007; Camp 2009). In our view, the complex relational abilities displayed by social creatures such as Tai chimpanzees qualify as concept-involving in that minimal sense. An anonymous reviewer points out that assigning the right label seems like a proto-linguistic ability that requires richer concept uses than those allowed in this minimal sense. We think that even if it can be regarded as proto-linguistic, this use still falls short of constituting a propositional activity.

2012 for a review). In this paradigm, the animal is set in a discrimination or memory test whose difficulty varies from trial to trial. At each trial, the animal has the possibility of carrying out the task, for which it can be rewarded when passing or not-being rewarded when failing, or avoiding doing it and receiving a consolation reward. These experiments aim at tapping the metacognitive capacities of the animal to the extent that animals can decide whether or not to carry out the task depending on the level of certainty that they have regarding their capacity to complete the task successfully. According to the paradigm, what distinguishes metacognitive regulation of behavior from mere regulation of behavior is that the selection of different behavioral schemas depends on information associated with the monitoring of the cognitive task, for instance, memory. So, the animal can regulate its decision to do the task or not depending on whether or not it is certain of its capacity to pursue it.

Now, our point is that Tai chimpanzees manifest similar metacognitive regulation of behavior when they can inhibit or execute different courses of action depending on the role they assume in cooperative hunting. As we mentioned, different roles are associated with different behavioral strategies and different cognitive skills like anticipating the prey's courses of action. However, the choice or inhibition of a particular behavioral schema depends on the updated information regarding the role that the chimpanzee assumes, which is dependent on different situational factors that go from the agent's position relative to the prey to the partner's behavior. That means that the agent constantly regulates its action schemata depending on the trajectory of the norms associated to the role in the cooperative strategy that they assume depending on the situation. So, it decides, for instance, whether to inhibit or not an attack schema depending on whether it recognizes itself as CAPTOR or not (Boesch 2002: 35–38). In other words, like the case of uncertainty, this is a case where the agent modifies its cognitive processes in accordance with information about itself. However, contrary to the case of uncertainty, the source of information is not a previous mental state (e.g., the animal's capacity). Instead, the information emerges from the conceptual normative expectation associated to the role that the animal assumes given the circumstance of the hunting.⁷

Now, one may object to our analysis that Tai Chimpanzees' hunting strategies do not require conceptual capabilities and that, in fact, a more deflationist reading of Boesch's findings is plausible.⁸ For instance, Tomasello et al. (2005) argue that humans exhibit a capacity for participating in collaborative projects with shared goals and intentions; however, in the case of Tai Chimpanzee's hunt, "there is nothing that would be called collaboration in the narrow sense of joint intentions and attention based on coordinated plans" (p. 685). Thus, without motivations to act as a team along with the lack of shared plans and goals, it is hard to see why chimpanzees' chase would require a conceptual repertoire that includes representation of (quasi) social roles like CHASER or AMBUSER. Moreover, such an argumentative move would not only cast into question the

⁷ It is worth mentioning that Tai Chimpanzees also exhibit the metacognitive capacity for selecting the role depending on internal information sources, for instance, their experience or expertise. As Boesch (200: 35–36) notices, the percentage of young chimpanzees with low expertise and low experience in hunting taking the roles of ambushers and blocks is significantly lower than those with higher experience, which implies that chimpanzees restrict themselves to assume certain roles depending on information they have about their expertise and experience.

⁸ We owe this objection to an anonymous reviewer.

conceptual nature of chimpanzees' metacognitive capacities in this context but also their social nature. In other words, one may argue that even if the Tai Chimpanzees hunting behavior reflects metacognitive capacities, they are not self-mindshaping capacities to the extent that they do not involve social expectations.

An answer to this objection can be partially found in Boesch's comment on Tomasello et al. (2005) paper itself. Boesch (2005) argues that the claim that chimpanzees lack pro-social motivation and the capacity to form shared plans and intentions is incongruent with several findings that report the propensity to hunt in collaborative groups (Boesch 1994; Stanford et al. 1994; Watts and Mitani 2000, 2002). In fact, Boesch argues that Tomasello's deflationist interpretation cannot accommodate several results of his studies. First, the roles taken by chimpanzees during the chase are complementary, which requires capacities for anticipation and coordination. Second, the roles can be shifted during the action, which necessitates at least a certain capacity for understanding the relationship between the different roles and the circumstances that make them applicable. These capacities are hard to explain without appealing to conceptual representations of the appropriate role and without some understanding of others' actions. Finally, the idea that the hunting strategies may be explained by selfish motivations is hard to make compatible with the fact that participants in the chase receive an equal amount of meat even when some roles like ambushers and drivers correlate with a low rate of success in capturing the prey.⁹ Thus, it seems difficult to explain cooperative hunting without postulating some cooperative inclination in chimpanzees along with the capacity for representing others' behavior as associated with a role in the strategy.

4.3 The role of propositions in metacognition

The third question on which accounts of metacognition may differ, following Proust's analysis, is concerned with the propositional character of metacognition. According to the attributive view, a necessary requirement for metacognition is the possession of mental concepts –such as believing, desiring, hoping, and so on– in order to gain theoretical knowledge about the role of our mental states and their content. When an agent acquires a mental concept, she acquires the capacity for metarepresenting a mental state, as in the canonical form 'S believes that *p*'. In contrast, the self-evaluative view regards metarepresentation as neither necessary nor sufficient for metacognition. First, it is not necessary because, according to Proust, there are cases where metacognitive capacities are displayed without metarepresentational capacities or even concepts, as the case of nonhuman animals would show. Second, it is not sufficient because, *prima facie*, we could metarepresent a mental state without evaluating its content, and it is the evaluative activity that is constitutive of metacognition. So even in analytic metacognition the role of metarepresentation is subordinated to the evaluation performed.¹⁰ In the previous section we argued, against Proust, that to explain the metacognitive abilities of animals we need to attribute them forms of

⁹ In Boesch's (2005) own words, "drivers and ambushers achieve only 1% and 11% of the captures respectively, while 81% are achieved by individuals following the hunt from the ground...Interestingly, ambushers that anticipate movements of the prey and the other hunters are granted an amount of meat equal to captors, even when they have not made the capture." (p. 692).

¹⁰ We will get back to this issue in section 5.2.

conceptual but nonpropositional control. Now we want to explain how the self-mindshaping view deals with propositional metacognition in a way that differs from the self-attributive view. In this section we will limit ourselves to two points: One, that there are cases in which metacognition is propositional but not metarepresentational; the other, that even in those cases in which metarepresentation is involved, it does not require mindreading abilities but mindshaping ones. We will leave for section 5.1 a further examination of the relation between self-mindshaping and metarepresentational metacognition.

First, given the range of mechanisms that are cited among mindshaping processes, it is plain that propositions play a prominent role in some of them, notably in narratives, which make conceptual connections in a coherent pattern. However, even if this is a kind of expertise that involves propositions, it does not necessarily depend on theoretical considerations about psychological constructs. In other words, there are cases in which one accesses one's mind propositionally yet without employing mental concepts, i.e., one can regulate one's cognitive processes by evaluating what follows from P or what is expected from P without considering one's attitude towards P. Deliberation offers an example of such a case in which a subject needs to monitor and evaluate her mental representations in order to reach an optimal decision. To this end she does not need to metarepresent her states in order to deploy metacognitive skills, she just need to access and regulate the contents of those states, which she typically does by means of linguistic vehicles. Moreover, those contents are not already there to be regulated, but the regulation consists precisely in giving shape (usually in linguistic form) to the content that best represents one's actual commitments towards the aspects relevant for the deliberation.¹¹

Consider the following example. Imagine I have to decide which of my favorite bands I will see tonight. Instead of making an intuitive decision I can decide to recruit some knowledge about my preferences (which live show do I prefer?) and beliefs (which, I think, is the best last album of each band?). This deliberative mode constitutes a form of metacognition (Mata et al. 2013) that facilitates the decision. Now, the recruitment of such meta-knowledge does not necessarily involve mindreading capacities or even capturing the appropriate information in meta-representational format. I do not read my mind to discover my innermost wishes. Rather, what I typically do is to deliberate about things such as which band is at its peak performance level, on which album they will draw for the set list, or which place has better acoustics. As Moran (2001: 60–65; see also McGeer 2004, pp. 246–247) argues, the process of recruitment is governed by *the principle of transparency*, according to which answering the question '*Do I believe that p?*' is analogous to answering the question '*Is p true?*' In order to gain knowledge about our own preferences and beliefs and to use them for regulating other cognitive processes, we do not quasi-perceive or infer our attitude toward a certain propositional content. In fact, more often than not, we make up our

¹¹ One may argue that, to the extent that deliberation requires a linguistic vehicle, it is necessarily a metarepresentational endeavor consisting in accessing the stable content represented in a linguistic format (Bermúdez 2003). However, we often engage in reasoning chains in inner speech format that lead us to a private judgement that we did not hold before (Fernández Castro 2019; see also Geurts 2018). Such kinds of reasoning demonstrate that we do not represent previous mental states, and that however, when we arrive at these judgments, we can modify or regulate our actions in accordance with the commitments associated to these judgments without the need of self-ascribing a mental state.

mind during the process of deliberation shaping our own preferences and beliefs when considering the consequences of how the world is in the light of the norms and values in place. Therefore, even though these metacognitive capacities require the use of propositional vehicles, they do not demand the metarepresentational format typical of mindreading.

Second, consider the use of narratives as an external model to regulate and interpret others' behavior according to certain standards. It is true that in these cases we resort to metarepresentational expressions of the type 'S believes that P'. However, this does not mean that we are reading the others' minds. From the mindshaping perspective, stories help to make better sense of people's behavior because they provide a coherent explanation of their responses to different circumstances. Now, for the purposes of metacognition, those narratives can be used as tools for regulating our own behavior as well. When we offer a coherent set of reasons for actions we are offering a coherent point of view that enables us to manifest our psychological stability by offering a self-narrative (Dennett 1992). Self-narratives generate expectations about our behavior, so that we respond to the environment in ways that live up to these expectations.¹² Velleman (2009) compares the self-shaping capacity of narratives with an actor improvising in a theatre. When the actor assumes the role of a character, he has to respond to the circumstances of the character on the spot, creating the role according to his personality, expectations, context, or the interaction with other characters. In the same vein, our behavior makes sense for us when it is coherent with our self-narrative, and we regulate it according to the expectations this narrative sets up, responding to others' behavior, personality and context. We are able to bring our personal narrative into line with our deeds and our deeds into line with our narratives.

Now, it could be objected that narratives do require mindreading skills. After all, narratives often include attributions of mental states to our past selves. However, there are reasons to doubt that these attributions actually rely on mindreading capacities. Arguably, a large part of our access to our past experiences and actions is subject to confabulation, As Dennett (1991, p. 94) puts it, people "fill in gaps, guess, speculate, mistake theorizing for observing". Empirical research suggests that confabulation is a pervasive characteristic of remembering and interpretation in our everyday life (Hirstein 2005). In confabulation we do not recall a past mental state: we make a story about a mental state that makes sense of our experiences in the light of rational norms. Once again, the idea is that one can resort to language as a vehicle for contents without committing to the notion that those contents are mental states that we find and that we insert in our current (metarepresentational) thoughts. Now, this can be extended to self-attributions of mental states in general: access to our past experiences is not a matter of metarepresenting past representations, but of generating them anew so as to bring coherence to a pattern of behavior by means of linguistic devices. What distinguishes recalling "real" states from confabulated ones is not that we have a privileged access to

¹² This view on the regulative role of narratives has to be distinguished from narrative theories of the self, which often endorse a stronger function for narratives as constitutive of self-identity (see Schechtman 2011 for a review). Our point is about a more "local" role for narratives as ways of controlling one's current mental states by aligning them with the expectations created by a certain storyline, which need not be a lifelong story but a more temporary contextually driven narrative. For a different formulation of a modest narrative self-shaping hypothesis see Hutto (2016).

the former: it is that we have better reasons to support them in a consistent story of our deeds and misdeeds.

To sum up, even if deliberating for a decision or telling ourselves about our attitudes require propositional competence, this is the effect of normative constraints of expectations and coherence, not of mindreading capacities. These skills are the intrapsychological counterparts of the interpsychological mindshaping capacities. Moreover, they are at the service of practical, non-epistemic goals that, as we will contend in 5.1, give self-mindshaping a certain primacy over self-attribution.

4.4 Metacognition as a practical mental action

The last question used by Proust to distinguish metacognitive theories regards the connection between metacognition and mental agency. According to the self-evaluative view—at least the view endorsed by Proust herself—the sense of agency, or the practical way we understand our mental states and processes as being subjected to intentional control, is connected to metacognitive capacities and the evaluation of these mental states and processes under epistemic norms. On the contrary, the self-attributive view contends that what we get when we access our (or others’) mental states is knowledge of our being in a certain state with a certain mental content. As the knowledge we gain and the action we take are different things, metacognition is not necessarily linked to mental agency.

As we said in section 2, the respective positions are also related to a difference regarding the states that fall under a metacognitive explanation. Suppose that I am seeing a flower or experiencing anger: According to the self-attributive view I can access these states and form the thoughts ‘I see a flower’ and ‘I am angry’. This is knowledge about myself that then I can put to different uses. As Proust puts it, in the self-attributive view “Thinkers can only evaluate their perception, their memory, their reasoning, and so forth, if they know that their present cognitive activity is one of perceiving, remembering, reasoning, (...) Attributing specific attitudes to oneself is a precondition for monitoring and controlling them” (2013, pp. 48–49). Proust rejects this requirement. According to the self-evaluative view, metacognitive abilities are involved in evaluating the accuracy of these perceptual and emotional states—e.g., whether I am seeing what I expected to see, or I am experiencing the appropriate emotion. To this end, it is not necessary to make a judgment, only to have an architecture in which each mental action comes with a nonconceptual monitoring mechanism that is capable of telling when the output deviates from certain standards.

Now, self-mindshaping agrees with the self-evaluative view on the point that metacognition is not a matter of judging what I am seeing or feeling, but a question of regulating these perceptual or affective states. Questions about what I am perceiving or experiencing arise in a certain context, and it is with respect to this context that metacognitive mindshaping takes place. For instance, establishing that I see a flower may arise in the context that I am smelling something different and I have to determine which of my senses is more reliable. Metacognition is related to my commitment to, say, my sight over my smell faculties. Thus a judgment such as ‘I see a flower’ or ‘I am angry’ is not a previous step, which accesses and metarepresents my current state so as to provoke a certain answer, but a result of my making up my mind about what I am actually perceiving or feeling. However, making one’s mind is also subject to the

normative force of how the world actually is –and “the world” includes elements such as that what I have in front of me is actually a flower, that my sight is normal, or that my autonomic nervous system is activated in a certain way.

Even if self-mindshaping agrees with self-evaluation on connecting metacognition with mental action, laying the emphasis on conformity and regulation, the question of agency is treated in a different way, closer to the concept-involving stance that underlies self-attribution. To be a rational agent is to be able to take responsibility of your own mental states by making them live up to the social norms that regulate our social interactions. This normative alignment can take place as a sort of automatic conformity that involves a sense of agency unrelated to knowledge of the content of the states so controlled. So it shares with the self-evaluative view the idea that there can be content-blind forms of control –even if the self-evaluative view does not include social conformity as one of those forms. However, there are other varieties of conformity that are not automatic and that demand an active self-commitment that constitutes a type of knowledge. The self-mindshaping view can distinguish between cases where our metacognitive processes run automatically and those that permit us to take a step back and adopt a deliberative stance (see Strijbos and De Bruin 2015, pp. 303–304). The latter seems to appear in cases that permit us to adopt a disengaged perspective to assess the mental states and processes and see whether or not they conform to certain norms. Yet, as we will see in the next section, this deliberative stance cannot provide a total disengagement from the practical concerns that characterize self-mindshaping. It is in this respect that one can defend a certain primacy of the latter with respect to self-attribution.

5 Is self-mindshaping complementary to other metacognitive processes?

Throughout this paper we have presented self-mindshaping as an account of metacognition that is distinct from both the self-evaluative and self-attributive views. In this section we want to consider whether it can be regarded as a complement rather than as an alternative to them. Given the variety of processes that can fall under the umbrella term, ‘metacognition’, it is easy to be tempted by a reasonable ecumenism in which the different views address different aspects of a multifarious phenomenon. Instead of yielding to this temptation, we want to explore to what extent self-mindshaping can be regarded as the central metacognitive phenomenon. Borrowing an expression from Peters (2019), we will call this the *primacy thesis* of self-mindshaping. First, we will examine it with respect to self-attribution, and then we will turn to self-evaluation.

5.1 The primacy of self-mindshaping over self-attribution

In order to assess the issue of the primacy of self-mindshaping, let us consider first this question from the perspective of the mindshaping vs. mindreading debate. The mindshaping view disputes the widespread idea that mindreading mechanisms are at the core of mental state attributions. However, as Peters (2019) points out, the relevance of mindshaping can be defended in different degrees of strength. In his view, prominent proponents of mindshaping, such as Zawidzki and McGeer, do not hold that

propositional attitude (PA) ascriptions only involve mindshaping. Instead, they are committed to what he calls the *primacy thesis*:

“PA ascriptions involve primarily a non-epistemic, practical capacity for mindshaping, not an epistemic capacity for tracking pre-existing PAs (e.g., the mindreading faculty), and any epistemic capacity for tracking pre-existing PAs is derived from and only as effective as it is because of (PA-specific) mindshaping, not *vice versa*” (2019, p. 539).

Peters advocates a weaker view in which “PA ascriptions involve and are for both mindreading and mindshaping equally” (2019, p. 538). Peters takes it that PA ascriptions, such as ‘Sue believes that P’, involve a commitment to the truth of a proposition, namely, the proposition $p = \text{Sue believes that P}$. This commitment demands an epistemic capacity to track evidence for p , i.e., a capacity for tracking pre-existing PAs. Moreover, this epistemic capacity often has the role of facilitating and strengthening mindshaping effects. In essence, this means that the primacy thesis is false: mindshaping and mindreading must be regarded as complementary to each other.

We think that Peters’s point about the facilitatory role of mindreading is not conclusive against the primacy thesis. As he explicitly acknowledges, mindshaping via PA ascriptions may occur without this commitment to truth and the epistemic capacities it involves. His point “is just that PA ascriptions are likely to have a *reduced* mindshaping impact” (2019, p. 547) in the absence of mindreading capacities. Now, if mindshaping is possible without mindreading but mindreading typically takes place to boost mindshaping, then one is entitled to hold that mindshaping is the primary process after all. In other words, one can defend the primacy thesis and not the complementarity view because the notion of complementarity suggests the existence of independent processes that complement each other, while the evidence alleged by Peters is about mindreading mechanisms tailored to subserve mindshaping’s goals.

However, Peters’s position can be reinforced by appealing to cases of PA ascriptions that occur with their own goals, and independently of mindshaping. The paradigmatic cases are those in which one adopts a *spectatorial stance* on PA ascription. Even though questioned by advocates of an interactivist approach as the main route to understand social cognition, it is undeniable that we can, and do often, engage in such a practice (Overgaard and Michael 2015). Consider a typical comedy of deceptions such as *Much Ado about Nothing*. From the perspective of its characters, the plot to instigate the belief that Hero is unfaithful to Claudio, or the centuries-old trick of the self-fulfilling prophecy that makes Benedick and Beatrice fall for each other, are paradigmatic cases of mindshaping at work. However, when I consider the comedy from my spectator’s point of view, I am engaged in an exercise of attribution of folk-psychological states to the characters, which does not require any mindshaping on my part. When I explain Claudio’s behavior on the basis of my belief attributions, the practical concerns that could lead to understand this explanation as a case of mindshaping are absent. As a spectator, one just wants to know about Claudio’s mental life for the sake of it, without any goals related to possible interactions, real or imaginary, with Claudio.

One might perhaps try to account for these attributions as a by-product of more fundamental mindshaping processes, even though it is unclear what kind of specifically mindshaping mechanisms could be at work here. What is important, at any rate, is that mindreading abilities can be deployed independently. Even if they derived from mindshaping, they acquire a life of their own – a life with its own epistemic goals, and its own processing resources. As the spectatorial stance towards other minds occupies a lot of our time, and not only while we watch fiction, it should not be dismissed as a marginal activity. The upshot is that mindshaping cannot be regarded as the primary mechanism behind PA in the sense posed by the primacy thesis. Mindreading abilities provide a complementary set of resources that can be employed to access other minds, sometimes in coalition with mindshaping, sometimes in a stand-alone manner.

Let us turn now to the question of central concern for this paper: can one hold a primacy thesis of self-mindshaping over self-attribution? That is, can one hold that “self-PA ascriptions involve primarily a non-epistemic, practical capacity for self-mindshaping, not an epistemic capacity for self-attribution”? Or are they complementary? At first blush, given what we just said about the putative independent deployment of mindreading mechanisms, it seems that the answer should lean towards complementarity. Since self-mindshaping and self-attribution do not posit mechanisms that differ from those employed in mindshaping and mindreading, any reasons invoked for the complementarity of the latter pair seem to militate on behalf of the complementarity of the former. However, the answer is not so straightforward. We just argued that an important motivation for complementarity lies in the fact that one can take a spectatorial stance toward other minds. The question now is: can one act as a spectator toward one’s own mind?

De Bruin (2016) contends that this is actually the case. In his view, one can take two stances toward oneself: a first-person stance in which one is engaged in making up her mind, and a third-person stance in which “you are able to step back from your first-person avowals, in order to reflect on them” (De Bruin 2016, p. 176). De Bruin contends that self-regulation involves a dynamic interplay between both perspectives, so mindshaping and mindreading turn to be complementary for our first-person folk-psychological capacities –i.e., they are complementary for metacognition. Notice, however, that this emphasis on the interplay is very similar to the facilitatory role of mindreading envisaged by Peters, and we argued that this is not enough to discard the thesis of the primacy of mindshaping. To this end *it is necessary to show that mindreading can be directed toward oneself in a spectatorial stance that is independent of any mindshaping attempt*. And this, we submit, is very problematic in first person PA ascriptions.

It is true that I can reflect on my own beliefs, ponder them, examine their basis, and so on. Yet this is not the kind of reflection that is driven by mere epistemic goals. The reason is that first person PA ascriptions do not allow the kind of disengagement that is typical of spectatorial instances. While in the case of third-person mental states one can act as an uninvolved anthropologist –who tries to make sense of the affairs of a human group without intending to affect, or be affected by, them– in the first person case one cannot but act as the engaged anthropologist –who seeks understanding as a way to contribute to solve some problems. While I examine my own attitudes, I cannot eschew my practical goals. I cannot conclude ‘well, I believe that P; this is interesting’ and then

carry on with my daily chores as if nothing had happened. When I conclude ‘I believe that P’ I inevitably transform my mind into one that is committed to my believing that P. This kind of reflection is a practical affair, not a purely epistemic one. It is not a matter of establishing truth-conditions but, as Peters (2019) puts it, of *forward-looking* truth-conditions, i.e., of what will be the case. Peters rejects that truth-conditions in the third person case are forward looking but he admits that they are plausible in the first person case “because in one’s own case, one has direct control over one’s own action and thinking” (2019, p. 545). Even when I ascribe myself a belief that I wish that would not be the case, the ascription is forward looking in the sense that I can take action to undermine this belief or to prevent it from getting control of my behavior. So first person ascriptions seem to be subsidiary to mindshaping.

To close this subsection, we want to consider a different way in which one might conclude that self-mindshaping is not the primary metacognitive process. The point is whether there are metacognitive functions that require mechanisms that are not constitutive of self-mindshaping. To elaborate on this point, let us begin by examining an analysis of the metacognitive abilities involved in the intentional control of the mind put forward by Vierkant (2012; also Vierkant and Paraskevaides 2012). Vierkant considers an example by Derek Parfit in *Reasons and Persons*. A rich young communist believes that most rich people become conservatives when they get older. He faces the dilemma of giving away the money according to his actual ideal or keeping it, and doing what he will consider right when he becomes a conservative in the future. Vierkant says that this case shows that “realizing that our first-order evaluations may change enables us to have a completely new level of self-control” (Vierkant 2012, p. 286). To exercise this control, you need to have knowledge of yourself as a psychological creature whose states can be manipulated. This scaffolds a range of metacognitive capacities, including those exercised by the young communist. He can understand that his mental states can vary, and thus, he can develop different behavioral and cognitive strategies in response. This capacity for flexible cognitive control requires understanding mental states as mental states, which in turn requires metarepresentation.

Notice that this position endorses a version of the primacy thesis stated above. First, it acknowledges that ascriptions involve primarily a non-epistemic, practical capacity. As Vierkant & Paraskevaides state: “the intentional acquisition of mental states will only get off the ground, if the subject evaluates that intentional control of the mental is the right thing to do” (2012, p. 122). It is clear that in Parfit’s scenario practical concerns are foremost for the rich young communist, who would not engage in the metacognitive intentional control necessary to solve his dilemma without the pressure to regulate his behavior to conform to the norms associated to the self-attribution of the pertinent mental states. Second, this kind of intentional control is derived from and only as effective as it is because of (PA-specific) mindshaping, since “when we intentionally manipulate our mentality, we do it on the basis of evaluations that have been shaped by social context” (2012, p. 122). In a nutshell, what metarepresentation provides is flexibility for a mind in which mindshaping is already at the heart of metacognitive regulation.

Nevertheless, the emphasis of this distinctive role for metarepresentational capacities raises the following question: where do those capacities come from? If they arise in a way that is independent from mindshaping mechanisms, then one could hold a

complementarity view between mindshaping and mindreading after all, undermining the primacy thesis. To answer that question what is at stake is what is meant by ‘metarepresentation’ in this context. Vierkant & Paraskevaides conceive a metarepresenter as “someone who understands that there are mental states and that those states are representational” (2012, p. 121). For instance, chimpanzees are not metarepresenters because even if they can understand themselves as *being* in a certain cognitive state –e.g., in the captor-related state– they have no insight about themselves as *having* cognitive states, so they can target the contents but not the states. In contrast, cases of intentional metacognitive regulation as the one exemplified by the rich communist, require having concepts of mental states.

So, how does one become a metarepresenter? Is it perhaps a linguistic ability, i.e., the capacity to manipulate certain symbols and to embed sentences within the scope of propositional attitude verbs? This is not the possibility favored by Vierkant, who distinguishes between the abilities provided by language –manipulation of representational objects– and the abilities provided by mastery of folk psychological concepts –flexible intentional control of mental states. The former metarepresentational resources do not seem to be enough to ground the latter metarepresentational capacities, so the question remains about the basis of these capacities. Here is where mainstream mindreading approaches could provide a strong alternative: the idea is that these metarepresentational capacities would be based on specific mechanisms to form and process metarepresentations. Examples of this can be found in Leslie’s influential account of a Theory of Mind Mechanism (Leslie 1987, 1994), which posits a mechanism to decouple a primary expression from its normal input-output relations, or in the hybrid theory/simulationist account proposed by Nichols and Stich (2003), which posits a special-purpose “Possible Worlds Box” that would construct representations of the world from another person’s perspective.

If one follows this route, it is clear that one has reasons to undermine the primacy thesis. Moreover, if you have to posit such strong mechanisms to account for mastery of mental concepts, why should you limit their use to the task of intentionally controlling your mental states? They could be employed for social cognitive and metacognitive purposes at large. However, we think there is an alternative way of explaining how one becomes a metarepresenter. The point is not to ground these metarepresentational capacities in specific mechanisms but in certain kinds of practices that are dependent on mindshaping processes. Acquiring mental concepts amounts to understanding the normative profile associated with the concepts, that is, the type of cognitive and behavioral ‘oughts’ we must draw from the given circumstances associated to the ascription of the concept. In other words, without being exposed to the types of practices and regulations postulated by the mindshaping view, one could not understand the normative structures that understanding mental concepts requires, and thus, one could not exercise the type of intentional control over one’s mental states. This way of understanding how to become a metarepresenter does not undermine the primacy thesis. In fact, even when one considers that the point about metarepresentation is to dissociate beliefs –qua psychological entities– from what one must do, considering what to do in the light on one’s future beliefs requires evaluating the normative profiles of commitments associated to the contents of the beliefs. The young communist is not only considering what his belief will be but the profile of the normative commitments associated to such a belief. Moreover, although

this kind of explanation seems to necessitate a certain notion of metarepresentation that allows the young rich communist to perceive himself as a minded creature, it does not commit us to understanding self-ascriptions as descriptions of inner processes or psychological states, rather than expressions that make explicit the commitments to the present and future actions associated with the content of the proposition (see Fernández Castro 2019 for a similar point).

Wrapping up: is it possible to maintain the primacy thesis of self-mindshaping over self-attribution? It seems that there are many cases—in particular, some that can be regarded as paradigmatically metacognitive—that need the goal-directed, normatively-constrained processes that regulate the formation of, and commitment to, one’s mental states. So to a certain extent, even if one is capable of pondering one’s attitudes in a quasi-detached manner, and even if one is capable of understanding oneself as a metarepresenter to better control one’s own mental processes, these capabilities often do not make sense in isolation from self-mindshaping activities. Now, this comes short of substantiating the primacy of self-mindshaping. To this end, it would be necessary to compare and evaluate specific models, testing specific hypotheses about the mechanisms actually involved in such and such metacognitive tasks. Doing this would go beyond the scope of this paper, yet we think that from our considerations we can sustain at least a moderate conclusion about the significant and independent contribution to metacognition that self-mindshaping processes make.

5.2 The primacy of self-mindshaping over self-evaluation

Before considering whether self-mindshaping and self-evaluation can be regarded as complementary or not, let us examine the issue of the relation between self-evaluation and self-attribution according to Proust (2013). As we said, she proposes a procedural architecture of metacognitive capacities where the mechanisms trigger different non-conceptual noetic feelings that indicate if the first-order processes or mental states are cognitively accurate. This is a metacognitive evaluative capacity that we share with certain nonhuman animals and that is still at work in adult humans (Proust 2015). But, as we noted in section 2, it is not the only metacognitive ability: alongside with it humans possess analytic metacognitive capacities that permit epistemic concept-based evaluations. These appear as a relatively late achievement (Kim et al. 2018), and they are largely based on the sort of mechanisms posited by the self-attribution view. However, Proust holds what can be regarded as a primacy thesis of self-evaluation over self-attribution. This thesis can be discerned in three aspects.

First, there is no complementarity between self-evaluation and self-attribution. This can be observed in her *exclusivist* view on metacognition, which is understood as “referring exclusively to the capacity of self-evaluating one’s own thinking” (2013, p. 313). Metacognition is a specialized capacity that “has a set of functional features of its own, which are independent of those associated with the self-attribution of mental states” (2013, p. 4). Second, both procedural and analytic metacognition are self-evaluative processes. So to the extent that some of the resources that are characteristic of self-attribution can be recruited in analytic metacognition, it is to subserve an evaluative capacity. Third, both the development and the performance of analytic metacognitive abilities are dependent on procedural metacognition. This is an important point, so let us quote Proust at length:

“analytic metacognition, although its verdicts may contradict feeling-based predictions, would not have been able to develop without an ability to intuitively appraise one’s certainty in selecting premises or in forming perceptual and memorial decisions. Second, analytic metacognition would not be able to come to a final decision about one’s certainty about a given proposal, if no metacognitive experience was available to stop the analytic regress to higher-order evaluations: even in its analytic forms, ‘feeling right’ puts an end to higher-order worries about normative appraisal” (2013, p. 70).¹³

Given the fact that, according to her, there can be procedural metacognition with no analytic metacognition, the resulting picture is very much a version of the primacy thesis for procedural metacognition over other metacognitive mechanisms.

Now it is time to examine the relation between the self-evaluative and self-mindshaping views. These views have an important aspect in common, which is their emphasis on norms. For Proust the capacity to evaluate or monitor mental actions depends upon certain norms, i.e., epistemic norms that allow one to discriminate good and bad performances and that set the threshold to trigger the correspondent feeling. In this respect, one could try perhaps to assimilate self-mindshaping to the kind of conceptual normative evaluation that Proust calls analytic metacognition. As we just saw that this metacognition relies on procedural metacognitive experiences, the upshot would be that self-mindshaping would rely on procedural metacognition too. Let us consider this issue by trying to extrapolate Proust’s considerations for the primacy of procedural metacognition in the development and performance of analytic metacognition.

First, does self-mindshaping develop from a procedural capacity based on noetic feelings? There are grounds to doubt this. Let us note that the actual function of noetic feelings, such as the feeling of uncertainty, is controversial. Carruthers suggests that such feelings have the evolutionary function of signaling “that the chances of successful action are low and to motivate actions that are designed to increase those chances” (Carruthers 2011, p. 291). As animals are interested in action in the world, “feelings of uncertainty (in both humans and animals) are more plausibly seen as directed at the world (in particular, at the primary options for action that are open to one), rather than at one’s own mental states” (Carruthers & Ritchie 2012, p. 82). The upshot is that feelings of uncertainty as such would not count as metacognitive. They appear as metacognitive in the human case because human uncertainty often contains metarepresentational concepts that result from our mindreading capacities (2011, p. 292). Now, we think that it is possible to accept Carruthers’s point about feelings being directed at action while at the same time eschewing the recourse to mindreading. Feelings can become metacognitive when the world, so to speak, “bounces back” to the mind in the form of an expectation. Feelings would allow people to become aware of their own expectations, hence turning them into a metacognitive state. Yet such feelings typically have a social cognitive origin. So even if the expectation-driven processes that characterize

¹³ This dependence is expressed in terms of dual-systems in Proust (2015, p. 22): “A purely automatic, reactive type of evaluation is possible, and is present in nonhumans and young children. It is prone, however, to generating throughout life illusions of competence and reasoning errors. A conceptually-controlled type of evaluation, on the other hand, can partially inhibit the influence of the expressive system, but it still depends on the latter to weigh the impact of context on ability, and to assess the trade-off between ease of processing and informativeness—that is, relevance—that is crucial in communication and in problem solving”.

mindshaping need “rightness and wrongness signals”, these signals would arise from social interactions in which others signal their approvals or disapprovals of my behavior. These signals can be subsequently internalized, in a Vigotskian-like manner, so that I can appraise by myself when my actions do or do not conform to certain expectations.

Second, does self-mindshaping face a regress that has to be stopped by appealing to rock-bottom metacognitive experiences? One may think that self-mindshaping needs the basic procedural architecture and its nonconceptual components (noetic feelings) in order to implement the conformity and pressure to fulfill the expectation generated by social norms. In other words, without noetic feelings that evaluate the output of the cognitive task, the agent could not fulfill the expectation generated by the norms. In order to reject this conclusion, let us consider two examples, one related to prospective noetic feelings, the other to retrospective ones.

Firstly, consider someone who is challenged to solve different math problems, so she has to evaluate her math competence and ask herself: “Can I solve these problems?”. Now, consider that the math problem comes from an elementary level textbook. The agent can generate normative expectations on the basis of what competence she should have according to her social status: “It’s a book for 10-years old, I *must* be able to solve this!”. These expectations are provided by the normative structure that dictates what kind of competence an adult should have given the social community in which she is embedded. This conceptual constraint is enough for metacognitive regulation, and prospective noetic feelings are not necessary for it. Indeed, without some conceptual guidance about what kind of problem I am facing, I can have no guidance from any internal feeling. I would be in the same position as someone who has to judge whether she can lift an object without knowing relevant information such as the material it is made of. One may have intuitive feelings about her competence, but without conceptual guidance, such feelings look gratuitous.

Secondly, against the necessity of retrospective noetic feelings consider the case of speech act evaluation. In principle, performing and understanding a speech act requires a plurality of sub-processes that need to be monitored, including conceptualization mechanisms, grammatical encoding, phonological encoding and so on. Now, given this plurality of processes it is quite implausible to consider that there are noetic feelings for all of them when they are working properly. Instead, it seems that we receive a warning only when something goes wrong, so retrospective noetic feelings mostly enter the scene in cases of failure. This notion can be extended to self-mindshaping processes in general. Success –e.g., in making a certain self-ascription– can often be accounted for by absence of “wrongness” feelings, which indicates that everything went as it should according to normative expectations. So self-mindshaping processes do not need procedural metacognitive experiences to do their duty.

We conclude, therefore, that there are no good reasons to hold a primacy thesis of procedural self-evaluation over self-mindshaping. What about the opposite thesis: can one hold a primacy thesis for self-mindshaping? Some of the considerations we just made point in this direction. On the one hand, even if “internal warnings” may sometimes have the form of noetic feelings, they derive from social mindshaping processes, such as the internalization of social signals, rather than be the basis for them. On the other hand, to be effective, and not mere ungrounded hunches, noetic feelings need to be supported by conceptual normative networks. However, we want to

be cautious about the generality of the primacy thesis of self-mindshaping over procedural mechanisms. One can leave the door open to some metacognitive role for noetic feelings without concepts in certain species. Our point is that when it comes to creatures with conceptual capacities, any metacognitive abilities are typically linked to conceptual mindshaping abilities.¹⁴ In sum, one can cast doubt on the idea that an architecture based on noetic feelings is necessary to account for cases of metacognition based on normative expectations. Even if it is true that sometimes normative expectations are associated with some feelings, they do not seem a necessary condition to account for the metacognitive access as conformity and regulation which the self-mindshaping view postulates.

6 Concluding remarks

As we have seen, the self-evaluative and self-attributive views of metacognition are not the only games in town. We have tried to motivate an alternative account of metacognitive capacities that derives from a particular understanding of social cognition as mindshaping. According to this approach, our practice of making sense of each other relies on a myriad of norms and routines that facilitate social coordination and interpretation. These normative structures are formed and maintained through different mechanisms, such as imitation, conformity to norms, or narrative production. We have argued that it is plausible to think that we can recruit some of these mechanisms for metacognitive purposes, that is, for shaping our own minds. This thesis lays the explanatory emphasis on mechanisms and processes that diverge from those at the center of the self-evaluative and self-attributive approaches.

Nevertheless, our main goal was not to speculate about specific details of these self-mindshaping mechanisms but to put on the table an alternative notion of metacognitive access. Instead of understanding thinking about thinking as a question of epistemic (knowing) or evaluative (controlling) access, we have emphasized the idea of metacognitive access as a practical normative notion. Moreover, we examined the prospects of a form of primacy of self-mindshaping processes over metarepresentational or procedural metacognition. We think that the prospects are good for the central cases, i.e., those that most people would regard as paradigmatically metacognitive. Those cases typically seem to involve practical goals and norms that drive and regulate our cognitive processes. Lending more support to this conclusion would require further examination of the merits of particular self-mindshaping accounts vis-à-vis alternative mindreading-based models. As the notion of mindshaping is relatively recent, it is clear the much work remains to be done to that effect. We are satisfied if we convinced the reader that it is at least an approach to metacognition that complements the extant alternatives, and that it is worth continuing to explore it.

¹⁴ An anonymous reviewer points out that we have not offered an argument to support the thesis that all instances of metacognition require socially instituted norms. So it is still an open question whether concept-using creatures can have a mindshaping-independent evaluative metacognition –i.e., some way of assessing the reliability of their cognitive processes that does not depend on social interaction. Even though we share Carruthers's reluctance to regard such assessments as metacognitive, we agree that his reasons for this rejection still come short as an argument against the possible coexistence of different (low and high) varieties of metacognition.

Acknowledgements This is a wholly collaborative paper, so order of authorship is arbitrary. The authors wish to thank two anonymous reviewers as well as the participants in the workshop 'Metacognition in individuals and groups' at the University of Stuttgart, where a preliminary version of this work was presented.

Funding Víctor Fernández-Castro's research is funded by Research Project ANR-16-CE33-0017, Agence Nationale de la Recherche, France.

Fernando Martínez-Manrique's research is funded by Research Project FFI2015-65953-P, Ministerio de Economía y Competitividad, Spain.

References

- Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965.
- Almagro Holgado, M. & Fernandez Castro, V. *Philosophia* (2019). <https://doi.org/10.1007/s11406-019-00096-2>
- Andrews, K. (2009). Understanding norms without a theory of mind. *Inquiry*, *52*(5), 433–448.
- Bernúdez, J. L. (2003). *Thinking without words*. Oxford: Oxford University Press.
- Boesch, C. (1994). Cooperative hunting in wild chimpanzees. *Animal Behaviour*, *48*(3), 653–667.
- Boesch, C. (2002). Cooperative hunting roles among Tai chimpanzees. *Human Nature*, *13*(4), 27–46.
- Boesch, C. (2005). Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behavioral and Brain Sciences*, *28*(5), 692–693.
- Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, *78*(2), 275–311.
- Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*(2), 121–182.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Carruthers, P., & Ritchie, J. B. (2012) The emergence of metacognition: Affect and uncertainty in animals. In M. Beran, J. Brandl, J. Perner & J. Proust (Eds.), *The foundations of metacognition* (pp. 76–92). Oxford: Oxford University Press.
- Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary Themes* (pp. 162–183). Cambridge: Cambridge University Press.
- Couchman, J. J., Beran, M. J., Coutinho, M. V. C., Boomer, J., & Smith, J. D. (2012). Evidence for animal metaminds. In M. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 21–35). Oxford: Oxford University Press.
- De Bruin, L. (2016). First-person folk psychology: Mindreading and Mindshaping. *Studia Philosophica Estonica*, *9*(1), 170–183.
- Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co..
- Dennett, D. C. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, & D. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp. 103–115). Hillsdale: Erlbaum.
- Evans, J. S. B. T., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.
- Fernández Castro, V. (2017). Inner speech in action. *Pragmatics and Cognition*, *23*(2), 238–258.
- Fernández Castro, V. (2019). Inner speech and metacognition: A commitment-based approach. *Logos and Episteme*, *10*(3), 245–261.
- Fodor, J. (1992). A theory of the child's theory of mind. *Cognition*, *44*(3), 283–296.
- Geurts, B. (2018). Making sense of self talk. *Review of Philosophy and Psychology*, *9*(2), 271–285.
- Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
- Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge: MIT Press.
- Hirstein, W. (2005). *Brain fiction: Self-deception and the riddle of confabulation*. Cambridge: MIT Press.
- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*(3), 164–181.
- Hutto, D. D. (2004). The limits of spectatorial folk psychology. *Mind & Language*, *19*(5), 548–573.
- Hutto, D. D. (2016). Narrative self-shaping. *Phenomenology and the Cognitive Sciences*, *15*(1), 21–41.

- Kim, S., Shahaecian, A., & Proust, J. (2018). Developmental diversity in mindreading and metacognition. In J. Proust & M. Fortiers (Eds.), *Metacognitive diversity: An interdisciplinary approach* (pp. 97–133). Oxford: Oxford University Press.
- Leslie, A. M. (1987). Pretense and representation: The origins of “theory of mind”. *Psychological Review*, *94*(4), 412–426.
- Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*, 211–238.
- Mameli, M. (2001). Mindreading, mindshaping, and evolution. *Biology and Philosophy*, *16*(5), 595–626.
- Martínez-Manrique, F. & Vicente, A. (2015). The activity view of inner speech. *Frontiers in Psychology* *6*
- Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373.
- McGeer, V. (2004). Autistic self-awareness. *Philosophy, Psychiatry, and Psychology*, *11*(3), 235–251.
- McGeer, V. (2007a). The regulative dimension of folk psychology. In D. D. Hutto, & M. Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 137–156). Dordrecht: Kluwer/Springer Press.
- McGeer, V. (2007b). The moral development of first-person authority. *European Journal of Philosophy*, *16*(1), 81–108.
- McGeer, V. (2015). Mind-making practices: The social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations*, *18*(2), 259–281.
- McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1–18.
- Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton: Princeton University Press.
- Nelson, K. (2007). *Young minds in social worlds*. Cambridge: Harvard University Press.
- Newen, A., & Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical Psychology*, *20*(3), 283–308.
- Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. New York: Oxford University Press.
- Overgaard, S., & Michael, J. (2015). The interactive turn in social cognition research: A critique. *Philosophical Psychology*, *28*(2), 160–183.
- 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.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge: MIT Press.
- Peters, U. (2019). The complementarity of mindshaping and mindreading. *Phenomenology and the Cognitive Sciences*, *18*(3), 533–549.
- Proust, J. (2001). A plea for mental acts. *Synthese*, *129*(1), 105–128.
- Proust, J. (2010). Metacognition. *Philosophy Compass*, *5*(11), 989–998.
- Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. Oxford: Oxford University Press.
- Proust, J. (2015). The representational structure of feelings. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*: 31(T) (pp. 1–25). Frankfurt am Main: MIND group.
- Rochat, P. (2002). Ego function of early imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 85–97). Cambridge: Cambridge University Press.
- Schechtman, M. (2011). The narrative self. In S. Gallagher (Ed.), *The Oxford handbook of the self* (pp. 394–416). Oxford: Oxford University Press.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*(3), 114–121.
- Stanford, C. B., Wallis, J., Mpongo, E., & Goodall, J. (1994). Hunting decisions in wild chimpanzees. *Behaviour*, *131*(1–2), 1–18.
- Sterelny, K. (2012). *The evolved apprentice*. Cambridge: MIT Press.
- Strijbos, D., & De Bruin, L. (2015). Self-interpretation as first-person mindshaping: Implications for confabulation research. *Ethical Theory and Moral Practice*, *18*(2), 297–307.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691.
- Tsai, J. C.-C., Sebanz, N., & Knoblich, G. (2011). The GROOP effect: Groups mimic group actions. *Cognition*, *118*(1), 135–140.
- Velleman, J. D. (2009). *How we get along*. Cambridge: Cambridge University Press.

- Vierkant, T. (2012). What metarepresentation is for. In M. J. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 279–288). Oxford: Oxford University Press.
- Vierkant, T., & Paraskevaides, A. (2012). Mindshaping and the intentional control of the mind. In F. Paglieri (Ed.), *Consciousness in interaction: The role of the natural and social context in shaping consciousness* (pp. 105–124). Amsterdam: John Benjamins.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Watts, D. P., & Mitani, J. C. (2000). Hunting behavior of chimpanzees at Ngogo, Kibale national park, Uganda. *International Journal of Primatology*, 23(1), 1–28.
- Watts, D., & Mitani, J. (2002). Hunting and meat sharing by chimpanzees of Ngogo, Kibale National Park, Uganda. In C. Boesch, G. Hohmann, & L. Marchant (Eds.), *Behavioural diversity in chimpanzees and bonobos* (pp. 244–255). Cambridge University Press.
- Zawidzki, 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.

Affiliations

Víctor Fernández-Castro¹ · Fernando Martínez-Manrique²

¹ LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

² Department of Philosophy I, University of Granada, Granada, Spain