Spontaneous Mindreading: A Problem for the Two-Systems Account

Evan Westra

(Forthcoming in *Synthese*)

Abstract: According to the two-systems account of mindreading, our mature perspective-taking abilities are subserved by two distinct mindreading systems: a fast but inflexible, "implicit" system, and a flexible but slow "explicit" one. However, the currently available evidence on adult perspective-taking does not support this account. Specifically, both Level-1 and Level-2 perspective-taking show a combination of efficiency and flexibility that is deeply inconsistent with the two-systems architecture. This inconsistency also turns out to have serious consequences for the two-systems framework as a whole, both as an account of our mature mindreading abilities and of the development of those abilities. What emerges from this critique is a conception of context-sensitive, spontaneous mindreading that may provide insight into how mindreading functions in complex social environments. This in turn offers a bulwark against skepticism about the role of mindreading in everyday social cognition.

1. Introduction:

For decades, social cognition research has been dominated by the idea that we navigate the social world by attributing mental states to other individuals in order to predict and explain their behavior – the ability known as "theory of mind" or "mindreading" (Carruthers 2013; Fodor 1992; Goldman 2006; Nichols and Stich 2003). This approach to social cognition has been quite fruitful, and has yielded an immense body of empirical knowledge about the development of our social cognitive abilities and their neural underpinnings (Baillargeon et al. 2010; Saxe and Kanwisher 2003; Wellman 2014). But philosophers and psychologists are nevertheless divided over how great a role these folk-psychological concepts actually play in our everyday lives. While many continue to assume that the mindreading paradigm is basically sound, others have suggested that it is deeply flawed as an account of our ordinary socio-cognitive abilities, and must be radically re-thought.

One of the most compelling skeptical arguments about mindreading draws our attention to the unbounded scope of paradigmatic folk-psychological inferences (Bermudez 2003; Morton 1996; Zawidzki 2013). This argument begins with the idea that belief-formation itself is a holistic, unbounded, "isotropic" process (cf. Fodor (1983)). Our actions can be informed by an indefinitely wide range of beliefs and desires. For instance, when I decide to take the metro rather than drive, I may do so because I believe that taking the metro is better for the environment, and I desire to make environmentally-friendly choices; it may also be because I

think that parking is expensive, and I wish to save money; or I may believe that I am being followed, and I wish to lose my pursuers on the crowded metro platform. There are, in other words, indefinitely many different folk-psychological ways to rationalize a particular action. When interpreting other people's actions, the argument goes, we are faced with the daunting task of sifting through this immense space of possible mental causes, and abductively inferring which belief-desire set best explains the action in question. Such a process would no doubt be incredibly demanding and effortful, and would place heavy demands on executive systems like working memory – that is, if the task were not completely intractable. It thus seems highly unlikely that we engage in this kind of inference during our everyday social interactions, which occur at a very rapid pace.

Motivated by these and other skeptical concerns, a number of theorists have proposed alternatives to mindreading in order to explain our everyday social-cognitive abilities. Some have suggested that we gain knowledge of mental states via automatic, perception-like processes (Gallagher 2008). Others have argued that we draw on folk-psychological narratives and social norms to predict behavior, rather than belief-desire inferences (Hutto 2012). Still others have suggested that many of our socio-cognitive abilities may be subserved by a combination of low-level associations between perceptions of behavior and domain-general attentional processes (Heyes 2014). It has even been suggested that these abilities are parts of dynamical systems that emerge during social interactions with multiple agents, and thus cannot be explained in individualistic terms and all (De Jaegher and Di Paolo 2007).

It is worth emphasizing the broad-reaching, often radical consequences of these anti-mentalistic proposals. Mindreading is widely believed to be central to many uniquely human social practices: linguistic communication (Grice 1991; Wilson and Sperber 2012), moral judgment (Mikhail 2007; Thomson 1976; Young et al. 2007), joint action (Bratman 1992; Tomasello et al. 2005) and establishing new social conventions (Lewis 1969). Our grasp of the psychological underpinnings of these activities hinges on the view that mindreading is a cornerstone of social cognition. If we abandon the mindreading paradigm, then our theories about these important human activities must also be re-thought.

In the context of this dispute over the scope of theory of mind in our everyday social lives, the two-systems account of mindreading (Apperly and Butterfill 2009; Apperly 2011; Butterfill and Apperly 2013) seems to offer something of a middle ground: on the one hand, it adheres to the

basic idea that some form of mindreading is pervasive in our everyday lives. But on the other hand, two-systems theorists agree with mindreading skeptics that the attribution of "full-blown" propositional attitudes such as beliefs is generally quite slow and effortful, places heavy demands on attention and working memory, and likely requires fairly advanced linguistic abilities. Because it is so demanding, proponents of the two-systems account agree that this form of reasoning is unlikely to contribute to many of the ordinary social practices that are often associated with it. When it does, it is likely scaffolded by some of the very processes proposed by anti-mindreading theories, such as social norms and narratives (Apperly 2011).

And yet, two-systems theorists also maintain that we are nevertheless equipped with an innately-channeled, automatic mindreading system that is constantly active in the presence of other agents. However, the representational capacities of this system, according to the two-systems account, fall well short of the kind of unconstrained belief-desire reasoning typically associated with mindreading. Instead, this "implicit" mindreading system is said to employ a limited set of quasi-mentalistic, mainly extensional concepts and inference rules that allow us to roughly predict behavior. Because of its limited representational capacities, this system exhibits a number of signature limits that distinguish it from genuine, "explicit" mindreading. Specifically, the implicit mindreading system is said to be insensitive to the fact that agents represent the world under a particular *mode of presentation*. For example, this system could never predict that Lois Lane would be surprised to see Clark Kent fly, even if she knew that Superman can fly, and that Superman is Clark Kent, because the implicit system would be insensitive to the fact that a single individual can be represented in a number of different ways.

Thus, according to this view, humans possess two "systems" for mindreading: the early-developing, automatic "implicit" system, and the later-developing, slow and effortful "explicit" system. Initially, human infants start out with just the implicit mindreading system, and as a result, their mindreading abilities are subject to its signature limits. As they get older, acquire language, and gain social experience, children develop explicit mindreading abilities, which start to emerge after their fourth birthday. Ultimately, these two systems exist side by side in adulthood, producing distinct types of mental state judgments in parallel to one another, creating a dissociation between implicit and explicit forms of mindreading (Low and Perner 2012).

The main goal of this paper is to offer a critique of the two-systems account of mindreading. Specifically, I will be arguing that the two-systems account is unable to accommodate the extant empirical evidence on one, very central type of mental-state attribution: the attribution of perceptual states or "perspective-taking." I'll further argue that these problems generalize to other aspects of theory of mind, and thus seriously undermine the two-systems account. What emerges in its place is a picture of mental-state attribution that lies somewhere in between the automatic, rigid information processing of the implicit mindreading system, and the slow, effortful reasoning of the explicit system. The evidence that I will discuss suggests that even "full blown" forms of mental-state attribution can be both fast and flexible, and that "implicit" forms of mindreading can be highly flexible and context-sensitive. This combination of speed and flexibility is achieved via the coordinated integration of domain-specific mindreading strategies with goals, attention and knowledge stored in long-term memory.

But while the main target of this paper is the two-systems account, this critique has broader implications for mindreading skeptics as well. This is because a key flaw in the two-systems account is that it accepts the skeptic's claim that genuine mindreading must be slow and cognitively effortful. A proper understanding of the underlying processes that enable mindreading shows that this is a mistake. Thus, the picture of mindreading that emerges from my critique of the two-systems account can also serve as a reply to the skeptic: we should not abandon the mindreading paradigm so quickly.

In the second section of this paper, I will discuss the general theoretical motivations for the two systems account. Then, in section 3, I will introduce the notion of perspective-taking as it occurs in the social cognition literature, and explain how the two-systems account purports to explain perspective-taking phenomena. In sections 4 and 5, I will argue that the evidence from perspective-taking undermines key claims about the implicit and explicit mindreading systems, respectively. In section 6, I will show how the problems from perspective-taking generalize, and ultimately undermine the two-systems account as a whole. In section 7, I'll discuss the fast, flexible conception of mindreading that emerges from my critique, and how it can serve as a bulwark against theory-of-mind skepticism.

2. Why two systems?

The motivation for proposing two systems for mindreading, its proponents argue, becomes especially clear when we consider the kinds of properties that human mindreading must possess

in order to successfully navigate ordinary social interactions. First, our mindreading abilities must be very fast and efficient, in order to keep up with the pace of ordinary behavior. Second, they must also be representationally flexible, since we need to be able to attribute an indefinite range of attitude contents to others in order to make sense of the complexity of human behavior. The problem, according to two-systems theorists, is that,

> [T]here is a tension between the requirement that mindreading be extremely flexible on the one hand, and fast and highly efficient on the other. Such characteristics tend not to co-occur in cognitive systems, because the very characteristics that make a cognitive process flexible – such as unrestricted access to the knowledge of the system – are the same characteristics that make cognitive processes slow and effortful. Instead, flexibility and efficiency tend to be traded against one another. This trade-off is reflected in Fodor's distinction between "modular" versus "central" cognitive processes. (Apperly 2013, pp. 73–74)

Thus, human beings need at least two mindreading systems because no single system could be *both* efficient and representationally flexible. According to this view, we rely on the fast and inflexible system when we need to rapidly anticipate what others will do, while we turn to the slow, flexible system when we need to carefully reflect on their specific beliefs. Thus, the reason the implicit system is unable to represent "full-blown" propositional attitudes is because these are thought to place heavy demands on working memory, which is slow but representationally flexible (Butterfill and Apperly 2013).[1] The implicit system gains its speed and efficiency from the fact that it can circumvent these forms of reasoning, and rely instead on a strictly limited set of quasi-psychological concepts and inference rules to automatically generate rough-and-ready predictions about behavior. But when accurate behavioral prediction means factoring in the *way* that an agent represents a particular state of the world, this system ought to make systematic errors. The explicit system, meanwhile, should be able to accommodate these cases; but this processing will inevitably be slow and effortful, and always goal-dependent.

---

[1] Why is representing "full-blown" propositional attitudes so demanding? Butterfill and Apperly write:

> On any standard view, propositional attitudes form complex causal structures, have arbitrarily nestable contents, interact with each other in uncodifiably complex ways and are individuated by their causal and normative roles in explaining thoughts and actions…. If anything should consume working memory and other scarce cognitive resources, it is surely representing states with this combination of properties. (Butterfill and Apperly 2013, pp. 609–610)

See Carruthers (2015c) for a critique of this argument.

The purported properties of the implicit mindreading system appear to derive from its modularity. In particular, Apperly (2010) emphasizes the essential role that *informational encapsulation* would play in the two-systems architecture. An informationally encapsulated, modular system could permit us to circumvent the need for effortful uses of working memory in many social interactions, thus rendering mental-state attribution fast and efficient, and even possible for young infants and non-human animals. However, such a system would also be representationally limited, due to its lack of access to working memory and stored knowledge. In other words, the tension between the need for flexible and efficient mindreading that motivates the two-systems proposal is explained by the trade-offs inherent in a modular, informationally encapsulated architecture.[2]

Moreover, Apperly (2010) suggests that informational encapsulation may provide part of the solution to the challenge raised by mindreading skeptics mentioned in the introduction. He argues that a modular, informationally encapsulated system could impose "hard constraints" on the scope of our folk-psychological inferences, thus limiting the need for complex, abductive reasoning. By restricting the range of possible inputs that it could process, and by sharply delimiting the kinds of inferences that could be made on the basis of those inputs, an encapsulated, implicit mindreading system offers us a way to render mental-state attribution computationally tractable. Thus, the notion of informational encapsulation seems to provide the two-systems account with both a basic architectural framework and a potent theoretical justification.

Problematically for the two-systems view, there is growing consensus among cognitive scientists that perceptual systems – the paradigms of modularity (Fodor 1983; Pylyshyn 1999) – are not, in fact, informationally encapsulated.[3] Instead, we find that abstract, conceptual knowledge "penetrates" even the earliest, most rapid stages of visual processing. For instance, there is evidence that feedback signals from inferotemporal conceptual areas impact processing in the visual cortex just 100ms following stimulus onset, well before the onset of endogenous attention (Wyatte et al. 2014). Similarly, Moshe Bar and colleagues have shown that conceptual

[2] There are a number of other well-known modularist approaches to theory of mind (Fodor 1992; Leslie et al. 2004; Scholl and Leslie 1999); however, these accounts tend not to sharply distinguish between implicit and explicit mindreading systems, as the two-systems theorists do. Although a discussion of these views is beyond the scope of this paper, it is likely that many of the arguments to come that are directed at the two-systems account will also pose challenges for them as well.

[3] For a recent review of this topic, see Ogilvie and Carruthers (2016).

information in the orbitofrontal cortex gets applied to rapidly transmitted, low spatial-frequency visual information, which is then projected back to mid-level and high-level visual processing areas 50ms before object-recognition takes place (Bar et al. 2006; Chaumon et al. 2014). There is also EEG evidence that linguistically encoded categorical distinctions (e.g. the lexical distinction between light and dark blue in modern Greek) can penetrate pre-attentional, pre-conscious processing in the visual cortex as early as 200ms after stimulus onset (Thierry et al. 2009). In short, there appear to be a number of pathways by which conceptual information stored in long-term memory can penetrate even paradigmatically modular systems, such as early visual processing.

The fact that vision is unencapsulated tells us something important: the trade-off between speed and representational flexibility is not mandated by our cognitive architecture.[4] But just because certain aspects of perceptual processing may be unencapsulated, it does not follow that there are no genuinely encapsulated systems. For instance, the analogue magnitude system, which served as a model for the implicit mindreading system (Apperly and Butterfill 2009), may well be impenetrable to goals and information stored in long-term memory (Feigenson et al. 2004). However, the question still arises: is fast, efficient mindreading truly informationally encapsulated, like analogue magnitude reasoning? Or is it more like early vision, and capable of using both top-down and bottom-up information to rapidly and flexibly interpret the social environment?

3. The case of perspective-taking:

A key test-case for the claim that the implicit mindreading system is truly encapsulated is the component of theory of mind known as "perspective-taking," which consists in the ability to represent what other agents see. In the empirical literature on the subject, there is a well-established distinction between two different "levels" of perspective-taking, which captures two different ways in which an organism might represent the visual perspective relation. "Level-1" perspective-taking consists in the ability to represent *what* another agent can see. Level-1 perspectives are construed as external, spatial relations that hold between agents and objects in their environments. This kind of relation depends primarily on environmental factors, such as

[4] In their own critique of the two-systems view, Christensen and Michael give a number of examples of well-studied cognitive systems that also succeed in achieving both flexibility and efficiency without the need for strong encapsulation, including the orbitofrontal cortex, the mid-level visual system, and language comprehension (Christensen and Michael 2015).

an unobstructed line-of-sight, lighting, and distance. To be a Level-1 perspective-taker thus consists in representing whether this external relation is present or absent, and forming appropriate expectations about behavior on this basis. For instance, a Level-1 perspective-taker would not expect an agent wearing a blindfold to reach towards a goal object in front of her, because the blindfold interrupts her line-of-sight.

"Level-2" perspective-taking, in contrast, appears to be uniquely human. It involves representing the *way* that other agents see the world. Rather than a direct relation between agents and their environments, the Level-2 perspective relation holds between agents and representational contents; however, it depends upon some of the same environmental factors as Level-1 perspective-taking, such as line-of-sight. The key difference between Level-1 and Level-2 perspective-taking is that only the latter is sensitive to the representational, aspectual nature of vision.

To illustrate, imagine that you and a partner are seated opposite one another at a table, and lying flat upon the table is a screen with the numeral "9" on it. In the purely extensional, Level-1 sense, you would both see the same thing: "9". But in the intensional, Level-2 sense, you would each see something different: while you would see the numeral *as* the number nine, your partner would see it *as* the number six. In other words, the more complex Level-2 relation permits us to track differences in mode of presentation.

In humans, Level-1 perspective-taking abilities emerge fairly early in development, and are even present in infancy (Luo and Johnson 2009; Masangkay et al. 1974; Moll and Tomasello 2006). A number of non-human animal species are also capable of Level-1 perspective-taking, including corvids, canines, and great apes (Bräuer et al. 2004; Bugnyar et al. 2016; Call and Tomasello 2008). The ability to represent Level-2 perspectives seems to emerge somewhat later in childhood, after the fourth year of life – the same age when children pass the standard false belief task. (Flavell et al. 1981; Low et al. 2014; Surtees et al. 2012). For this reason, Level-2 perspective-taking is said to signal children's acquisition of a representational theory of mind (Rakoczy 2015).

Beyond its comparative and developmental applications, the Level-1/Level-2 distinction has also been invoked to describe adults' perspective-taking abilities. Specifically, it has been argued that representing Level-1 and Level-2 perspectives involve distinct cognitive processes (Michelon and Zacks 2006; Qureshi et al. 2010; Samson et al. 2010; Surtees et al. 2012; Surtees,

Samson, et al. 2016). Level-1 perspective-taking appears to be very rapid, places relatively few demands on executive resources, and seems to employ a simple line-of-sight heuristic. Level-2 perspective-taking, in contrast, appears to be slow, places heavy demands on working memory, and employs a kind of embodied mental rotation procedure (Surtees et al. 2013a).

The distinction between Level-1 and Level-2 perspective-taking thus seems to offer a clear-cut case of the dissociation between the implicit and explicit mindreading: Level-1 and Level-2 perspective-taking each possess developmental and cognitive profiles that map fairly neatly onto the two mindreading systems. Accordingly, the two-systems account makes a number of specific predictions about perspective-taking that bear directly upon the issue of informational encapsulation. First, if the implicit system is truly informationally encapsulated, then Level-1 perspective-taking should be insensitive to the background knowledge of the perspective-taker. Second, if Level-2 perspective-taking truly places heavy demands on working memory, then we should expect it to operate in a goal-dependent fashion, and to be relatively slow and effortful.

To test the first prediction, Samson et al. (2010) created the "dot-perspective task." In this task, adult participants had to rapidly judge what either they or an avatar could see. Subjects were presented with a scene in which an avatar stood alone in a room facing a wall. In Consistent Perspective trials, black dots appeared on the wall that the avatar could see. In Inconsistent Perspective trials, some of the dots appeared on the wall that the participant could see, but the avatar could not. In the Self-task, participants had to judge how many dots they themselves could see; in the Other-task, they had to judge how many dots the avatar could see. Samson and colleagues found that people were much slower to respond and made more errors in the Self-task for Inconsistent perspective trials. Participants seemed to represent the avatar's Level-1 perspective even when it was irrelevant to their current goal, to the point that it interfered with their performance – exactly as the two-systems account predicted it would.

To test the prediction that the implicit system cannot represent Level-2 perspectives, Surtees et al. (2012) presented participants with another scene containing an avatar; but this time, instead of dots, the experiment used numerals displayed on a table in front of the avatar opposite the participant – the "number-perspective task." On Consistent Perspective trials, a numeral like '8' was displayed, which both the avatar and the participant saw the same way. In Inconsistent Perspective trials, a '6' or a '9' was presented on the table, which the participant and avatar would perceive differently; thus, this task required Level-2 perspective-taking abilities. As in

Samson et al., participants completed both Self and Other tasks. Unlike in the Samson et al. experiments, the Inconsistent perspective of the avatar did not interfere with their response times on the Self-task. Participants appeared to only compute the other individual's perspective on the Other-task, when it was goal-relevant – once again, just as the two-systems account predicted.

While these results do seem to bear out the above predictions, a number of other findings in the perspective-taking literature are not so easily accommodated by the two-systems framework. In the next section, I will argue that we have good evidence that Level-1 perspective-taking is neither fully encapsulated nor truly automatic. In section 5, I will argue that Level-2 perspective-taking need not be slow and cognitively effortful.

4. Level-1 perspective-taking is unencapsulated: The argument from gaze-cueing

To see why the Level-1 perspective-taking evidence does not fully support the two-systems account, we need to consider another experimental paradigm that also aims to study implicit perspective-taking: gaze-cueing. Gaze-cueing tasks measure the effects of shifts in the direction of a target's eye gaze or head on covert spatial attention – that is, changes in attention that happen *prior* to any overt forms of attention shifting, such as movements of the eyes or head (Posner 1980). In gaze-cueing studies, subjects are presented with a task-irrelevant face in the center of a screen, with eyes that move either in one direction or another (Friesen and Kingstone 1998; Hood et al. 1998). Subjects then witness an object suddenly appear either on the same side as the direction that the face's eyes have "looked" (a congruent trial) or on the opposite side (an incongruent trial). The gaze-cueing effect occurs when subjects are faster to detect the object on the congruent side than the incongruent one. These effects are extremely rapid – on the order of 10-15ms - and are also specific to social stimuli (Kingstone et al. 2004; Ristic and Kingstone 2005).[5] Thus, gaze-cueing seems like exactly the kind of effect that one

[5] Since cueing effects can also be triggered by other kinds of directional stimuli, such as arrows (Ristic et al. 2002), some have suggested that this process might be the product of a domain-general covert orienting mechanism (Santiesteban et al. 2014). However, these two types of cueing effects appear to have distinct cognitive, developmental, and neural bases. Specifically, gaze shifts appear to issue in a distinctly spatial cueing effect for the specific location where the eyes look, whereas arrows produce object-based cueing effects for any items that appear on the congruent side, regardless of their specific location (Marotta et al. 2012). Further, while gaze-cueing effects appear even in extremely young infants (Farroni et al. 2009; Hood et al. 1998), cueing effects from other kinds of stimuli do not emerge until much later in development (Jakobsen et al. 2013). Finally, gaze-cueing, but not other kinds of cueing, produces activity in the superior temporal sulcus (STS), a neural region associated with social cognition (Ristic and Kingstone 2005) (see also Michael and D'Ausilio (2015).

might expect from the implicit mindreading system: it is extremely fast, unconscious, and tracks Level-1 perspectives.

If the implicit system were truly encapsulated, knowledge stored in long-term memory would not affect it. However, we know from a wide range of studies that gaze-cueing is in fact sensitive to background knowledge. For instance, Eva Wiese and colleagues showed participants a robot-face cueing stimulus (Wiese et al. 2012). In one experiment, they found that participants were much less likely to be cued by the gaze-shifts of the robot than those of a human face. However, in another experiment, participants were explicitly told that an experimenter was intentionally controlling the robot's gaze-shifts. In this condition, participants were just as likely to be cued by the robot-face as the human face. Thus, the presence of explicit, folk-psychological background knowledge about the stimulus affected whether or not partners were cued by an otherwise non-agentive stimulus.

Similarly, when a cueing stimulus is ambiguous, background knowledge about whether or not it is an intentional agent can modulate whether it produces a cueing effect. Ristic & Kingstone (2005) showed subjects an ambiguous stimulus, and told them that two eye-like shapes were either eyes or wheels on a car; they found cueing effects for the eyes condition, but not for the car condition. Even more strikingly, Terrizzi and Beier[6] recently showed participants an unfamiliar entity and modulated whether or not, prior to the cueing trials, subjects saw another agent appear to interact with it in a contingent, seemingly social manner. They observed "gaze" cueing effects for the unfamiliar entity (even though it did not, in fact, possess eyes, but merely a presumed front and back) in the social interaction condition, but not in the non-social condition.

Background knowledge about whether or not a human face can see also modulates the cueing effect. Teufel and colleagues showed participants images of a face wearing goggles; beforehand, subjects had the opportunity to handle a seemingly identical pair of goggles (Teufel et al. 2010). However, one group handled goggles with opaque lenses (such that the wearer would not be able to see through them), while another group handled goggles with transparent lenses. They found that only participants who handled the transparent goggles were cued by the head-

[6] Submitted manuscript: "Automatic Attentional Cueing by a Novel Agent in Preschool-Aged Children and Adults" (personal communication).

movements of the stimulus. Thus, if participants knew that the face could not see, the cueing effect was attenuated.

Importantly, these studies always showed subjects in both experimental and control conditions *perceptually identical stimuli*; all they varied was the background knowledge that subjects had about what they were looking at. In other words, these studies provided a perfect test for informational encapsulation, and showed that gaze-cueing is not encapsulated after all. Thus, contrary to the two-systems account, background knowledge affects Level-1 perspective-taking.

One study from the two-systems group offers a potential avenue for them to respond to this point. Using the same stimuli as in the Samson et al. study described above, Qureshi et al. (2010) tested whether or not Level-1 perspective-taking would be affected by concurrent executive demands; according to the two-systems account, it should not. To do this, they used a dual-task interference design in which subjects had to complete the dot-perspective task while simultaneously tapping along with a recorded beat. They found that the cognitive load task did interfere with task performance, but this interference was similar for both the Self- and Other-tasks. While this finding might initially be interpreted as undermining the claim that Level-1 perspective-taking is truly an efficient process, the authors argued that the similar interference effects for both Self- and Other-trials showed that the tapping task did not interfere with the *calculation* of Level-1 perspectives as such, but rather with the attentional *selection* of perspectives in general. According to this picture, the Level-1 perspective-taking process would involve two components: a perspective-selection process that places demands on domain-general attention, and a domain-specific, encapsulated mechanism for perspective-calculation.

Accordingly, proponents of the two-systems account could argue that all the gaze-cueing studies show is that the Level-1 perspective-*selection* process is unencapsulated from background knowledge, but that the perspective-calculation process is not. Thus, in cases when the gaze of a target face is known not to be indicative of genuine seeing, that perspective might not be selected by attention, and thus no perspective-calculation would occur. But it could still be maintained that Level-1 perspective-*calculation* is encapsulated, once a given perspective has been selected.

This distinction between selection and calculation enables the two-systems theorist to maintain that that there could be an encapsulated mechanism for Level-1 perspective-calculation. But at

best, such a mechanism could only be one component of the system that performs the function of Level-1 perspective-taking. This is because perspective-selection also seems to be a necessary part of the perspective-taking process: in the absence of perspective-selection, no perspective-taking could take place. Thus, we could not ascribe the function of Level-1 perspective-taking solely to the perspective-calculation mechanism. If there is a "system" that is responsible for Level-1 perspective-taking, then it must also include whatever mechanism or mechanisms that accomplish perspective-selection – and these, it appears, are unencapsulated. Thus, while the "system" responsible for Level-1 perspective-taking might involve component parts that are informationally encapsulated, this does not change the fact that Level-1 perspective-taking as such *is* sensitive to background knowledge.

Moreover, acknowledging a role for domain-general attention in the perspective-taking process also undermines the claim that Level-1 perspective-taking is truly *automatic* – that is, if by "automatic" we mean a process that is mandatory, stimulus-driven, and goal-independent (Moors and De Houwer 2006). This is because, more often than not, domain-general attention is goal-directed (Carruthers 2015a). In paradigmatic instances of "top-down" attentional orienting driven by the dorsal orienting network, these goals are conscious. But attention can also be controlled by the ventral orienting network, which is sensitive to unconscious goals and motivations (Corbetta et al. 2008).[7] Thus, by acknowledging a role for attention in perspective-selection, two-systems theorists are opening up a space where goals might play a significant role in the Level-1 perspective-taking system.

Consistent with this possibility, other studies have shown that knowledge of the social group memberships of a target face, including its age, race, social status, and perceived threat all affect gaze-cueing (Chen and Zhao 2015; Dalmaso et al. 2012; Pavan et al. 2011; Slessor et al. 2010). In addition to interactions between the gaze-cueing mechanisms and long-term memory, these findings show that gaze-cueing is also sensitive to motivational factors: when a face is motivationally salient – for instance, because it belongs to a threatening out-group member – we preferentially allocate attentional resources in order to follow its gaze. However, when a

[7] Granted, attention can sometimes be "captured" in an automatic, goal-independent manner by environmental stimuli (Knudsen 2011), and it's conceivable that Level-1 perspective-taking could likewise be the product of purely bottom-up processing. However, many of the gaze-cueing experiments cited above were able to perfectly control for such low-level effects by using perceptually identical stimuli in both experimental and control conditions. The factors that modulated Level-1 perspective taking in these experiments could not have been purely stimulus-driven.

face is not motivationally salient – say, because it belongs to a low-status in-group member – we do not preferentially attend to its gaze direction. In other words, Level-1 perspective-taking appears to be highly sensitive to our social goals.

Thus, the evidence from gaze-cueing seems to shows that Level-1 perspective-taking is neither wholly encapsulated, nor truly automatic. Of course, Level-1 perspective-taking is also not under explicit, top-down, conscious control. Rather, its information-processing profile seems to belong somewhere in between these two. It is better described as a "spontaneous" process: it is fast, efficient, and unconscious, but also sensitive to background knowledge and goals (Carruthers 2015b). Notably, this kind of process does not quite fit with the descriptions of either the implicit or explicit systems. Instead, it seems to share attributes of both.

If this picture is right, and Level-1 perspective-taking is really spontaneous, rather than automatic, then why do subjects in the dot-perspective task represent the avatar's Level-1 perspective? This did, after all, conflict with their overt goal, and it is not obvious what else might have motivated participants to attend to its perspective. One possibility is that even though Level-1 perspective-taking is not genuinely automatic, we may possess a standing disposition to represent other agents' perspectives when doing so is cognitively efficient.[8] Given that what other agents can see tends to be behaviorally relevant, and that calculating Level-1 perspectives is not particularly demanding, such a disposition would be fairly adaptive in most situations. In practice, this might make Level-1 perspective-taking seem automatic in most situations, when in fact it is really motivation-dependent.

5. Level-2 perspective-taking can be fast and efficient

The argument from gaze-cueing suggests that Level-1 perspective-taking does not quite fit with the description of the implicit mindreading system as automatic and encapsulated. However, it leaves untouched the basic claim that Level-2 perspective-taking should be a slow, effortful, working-memory-based process. Thus, two-systems theorists may be willing to concede that Level-1 perspective-taking is more flexible than they initially supposed, but still argue that Level-2 perspective-taking, which involves "full-blown" propositional attitude

[8] Along similar lines, Fiebich & Coltheart (2015) suggest that which socio-cognitive procedure we use is determined by whether or not it will be cognitively effortful in a given context (Fiebich and Coltheart 2015). (Thanks to an anonymous reviewer for bringing this reference to my attention).

attribution, must possess something like the cognitive profile of the explicit mindreading system.

Notably, Level-2 perspective-taking tasks almost always involve some kind of mental rotation (Flavell et al. 1981; Low et al. 2014; Surtees et al. 2013b), as this seems to be one of the most straightforward empirical methods for creating a dissociation between Level-1 and Level-2 perspectives. Problematically, mental rotation is known to place heavy demands on working memory even when mental-state attribution is not involved (Hyun and Luck 2007). Peter Carruthers has recently argued that this role for mental rotation constitutes a serious confound for many Level-2 perspective-taking tasks, and that these tasks do not so much demonstrate a difference in the concepts of *seeing* deployed in Level-1 and Level-2 scenarios or a difference in underlying mindreading systems as a difference in non-mentalistic task demands (Carruthers 2015b, 2015c). As an alternative explanation, Carruthers suggests the lack of altercentric interference in the number-perspective task was due to motivational factors: because they were not sufficiently motivated to represent the avatar's perspective, subjects in this task simply did not go to the trouble of mentally rotating the numeral on the table.[9]

One interesting possibility that emerges from Carruthers' motivation-based interpretation is that changing the motivational structure of the number-perspective task could potentially lead participants to maintain a representation of the other agent's Level-2 perspective. Elekes and colleagues investigated this possibility by creating a modified version of the number-perspective task, which subjects either completed by themselves (the Individual condition) or with another participant (the Joint condition) (Elekes et al. 2016). This initial modification of the number-perspective task is especially noteworthy: while a nondescript avatar might be salient enough to warrant Level-1 perspective-taking, it is not obvious that participants would care enough to go to the trouble of maintaining a representation of its Level-2 perspectives. Exchanging the avatar for a live human being both increases the potential salience of the target (real people are generally more interesting than nondescript avatars), and improves the ecological validity of the paradigm. As we'll see shortly, this manipulation proves to be effective.

[9] Carruthers does accept that the evidence from the dot-perspective task shows that Level-1 perspective-taking is automatic, although he denies that these results are best explained in terms of a non-representational concept of seeing. On his "one-system" account, the attribution of mental state concepts is automatic when executive resources are not required, and "spontaneous" when they are. However, the argument from gaze-cueing from the previous section shows that even Level-1 perspective-taking is a spontaneous activity, rather than truly automatic.

Subjects in this experiment completed a number-verification task, which involved rapidly judging whether the number they saw on a screen lying flat in front of them was the same as the number they heard in an audio recording. In the Joint condition, experimenters manipulated whether participants believed that the person seated across from them was completing the same number-verification task (the perspective-dependent task), or an *n*-back task in which subjects had to judge whether or not the color of the number on the screen was the same as the number that came before it (the non-perspective-dependent task). Thus, in both tasks in the Joint condition, subjects knew that their partner was also attending to the numeral on the screen, but only subjects completing the perspective-dependent task believed that their partner was attending to the same aspects of the numeral (namely, its value). But importantly, all subjects ever had to do was complete their own task; their partner's performance was irrelevant.

The experimenters found that subjects in the Joint condition were slower than in the Individual condition, but only when both completed the perspective-dependent task *and* the numerals of the screen were such that their values differed on the basis of perspective (i.e. 2, 5, 6 and 9); for numerals whose values appeared to be the same regardless of which side of the table the participant was at (i.e. 0 and 8), there was no difference between the individual and joint conditions. In effect, subjects were only slower when 1) they had a live partner, 2) they believed that their partner had a similar goal, and 3) the partner's response would diverge from their own on the basis of their Level-2 perspective. These results suggest that knowing that a partner possesses a similar goal to one's own creates an unconscious motivation to maintain a representation of their perspective, even when this is not relevant to one's overt goal. When this representation differs from one's own first-personal one, this creates altercentric interference.

Using a very similar design, Surtees and colleagues obtained a slightly different set of effects (Surtees, Apperly, et al. 2016). Like Elekes et al. (2016), they used a number-verification task that used live partners seated on opposite sides of a display that lay flat on the table between them; in one of the experiments, Surtees et al. also included a version of that task where one partner made judgments about a surface feature of the numeral on the screen, rather than its value. And just like in the Elekes et al. design, subjects only ever had to judge the value of the number from their own perspective – the perspective of the other participant was always task-irrelevant. However, in the Surtees et al. (2016) design, subjects took turns instead of

completing the task at the same time; turn-taking either occurred within the same block of trials (with the two participants alternating rapidly), or in separate blocks (with one participant going first and the other going second).

Like Elekes and colleagues, Surtees et al. found that the presence of a live participant affected subjects' Level-2 perspective-taking, with an altercentric interference effect when their perspectives were inconsistent, and also a facilitation effect when their perspectives were the same. But unlike Elekes et al., they found that altercentric interference arose even when the partner was attending to surface features of the numeral, rather than its value. They also found that altercentric inference did not occur in subjects who went first when completing the task in separate blocks; however, when the second partner took her turn, the altercentric interference effect re-emerged.

Collectively, the results of Elekes et al. (2016) and Surtees et al. (2016) yield a number of conclusions regarding Level-2 perspective-taking, as well as some open questions. First, using a live participant instead of an avatar seems to increase the likelihood that subjects will spontaneously adopt another agent's Level-2 perspective, even when it is not relevant to their overt goals; however, the mere presence of a live participant is not sufficient for this to occur. In the simultaneous task design of Elekes and colleagues, participants only took their partner's perspective into account when explicitly informed that they were performing the same task. In the turn-taking design of Surtees et al., subjects only adopted their partner's perspective when they had previously observed their partner completing the task that they themselves were about to undertake. In both cases, some form of prior knowledge was necessary for spontaneous Level-2 perspective-taking to occur.

The fact that subjects in the Surtees et al. task spontaneously adopted the perspective of their partner even when the partner was not attending to a perspective-dependent feature of the numeral on the screen is inconsistent with the findings of Elekes et al. However, this difference may be due to the difference between the alternating turn-taking task design used in the former study, and the simultaneous task design used in the latter. It is possible that the turn-taking activity created the sense of a shared goal, when in fact there was none.

The most important conclusion to be drawn from this set of findings is that Level-2 perspective-taking can, at times, be fast and efficient, provided that subjects are provided with the right background knowledge and are sufficiently motivated. This contradicts the claim that

Level-2 perspective-taking is a slow and effortful process. In more ecologically valid tasks that use a live participant rather than an avatar, Level-2 perspective-taking turns out to be spontaneous (just like Level-1 perspective-taking).

These findings create something of a puzzle for both the two-systems theorists and its critics, such as Carruthers: if Level-2 perspective-taking tasks place inherent demands on working memory (either because working memory is a constitutive part of explicit mindreading more generally, or because of the mental rotation confound), how come subjects were able to efficiently generate Level-2 perspective representations in these circumstances? The answer may be related to the fact that spontaneous perspective-taking only occurred when subjects possessed the appropriate prior knowledge (in addition to the right motivations). Once subjects learned that their partners' perspective systematically differed from their own (e.g. "If I see 6, he sees 9"), they would have been able to store that knowledge as a mentalistic schema in long-term memory, where it would have been available for rapid retrieval.[10] Thus, even if subjects had to initially engage in effortful mental rotation to judge their partner's perspective, they would subsequently be able to infer their perspective without any effortful spatial reasoning at all. By using memory-based strategies, subjects would have been able to circumvent the need for any effortful use of working memory.[11]

It is worth noting that Apperly (2010) does discuss one possible way that explicit, demanding forms of mindreading could be rendered fast and efficient: *downwards modularization.* The basic idea behind downwards modularization is that expertise can render otherwise demanding tasks fast and efficient. For example, where an average chess player might discover a path to checkmate through slow, effortful reasoning, an expert player might, thanks to her extensive experience, arrive at a similar conclusion in a seemingly effortless manner. One way that this sort of efficiency-through-expertise can be achieved is when a body of knowledge – initially acquired through explicit, effortful processes – is used so often that it leads to the formulation of cognitive schemas. These schemas enable us to rapidly pair inputs to the appropriate

[10] Christensen and Michael (2015) discuss the use of schemas in mindreading at length in their "cooperative multi-systems architecture" proposal, which they offer as an alternative to the two-systems account.

[11] Interestingly, Michelon and Zacks discovered that subjects also tended to use memory-based strategies in a Level-1 perspective-taking task: instead of calculating the line-of-sight of an agent directly, participants simply memorized the set of objects that the agent could see, and this led to increased performance (Michelon and Zacks 2006). The experimenters, who were interested in studying how line-of-sight is calculated, developed a method to control for this strategy. But it highlights the fact that memory-based perspective-taking strategies provide an ever-present, efficient alternative to the use of more spatial forms of reasoning, whether these involve line-of-sight calculation or mental rotation.

behavioral outputs without having to go through any effortful, explicit reasoning. But, according to proponents of downwards modularization, this efficiency is achieved at the cost of flexibility. Just like innate "original" modules, these acquired modules are ultimately encapsulated from goals and background knowledge. Apperly suggests that downwards modularization might often occur with our explicit mindreading abilities: an expert poker player may, for example, become so well-practiced that she is able to automatically detect a bluff without needing to engage in any explicit reasoning at all.[12]

However, the effects on Level-2 perspective-taking described above are not plausibly the result of downward modularization. First, subjects never had the explicit goal of monitoring the other agent's perspective at all; Level-2 perspective-taking was actually detrimental to their performance on the explicit task. Expertise, in this context, would consist in ignoring the partner, not representing the way she saw the number. Second, it is implausible that subjects came into the experiment with an acquired module for Level-2 perspective-taking. If this were the case, then altercentric interference should have been present across all the Joint conditions (or, in the case of the Surtees et al. findings, the conditions where partners were merely present, but not yet engaged in the number-verification task), not just the ones where subjects shared a similar goal. The fact that these altercentric interference effects were so context-sensitive suggests that the Level-2 perspective-taking abilities brought by subjects to the lab were flexible and goal-dependent, not stimulus-driven. Thus, the fast and efficient Level-2 perspective-taking that we find in these studies seems to occur in spite of the fact that it is unencapsulated, which runs contrary to the downwards modularization proposal.

6. Implications for the two-systems account

The arguments of the last two sections create serious problems for the two-systems account of perspective-taking. Contrary to that framework, it appears as though both Level-1 and Level-2 perspective-taking can be fast and efficient, but also sensitive to goals and background knowledge. Thus, both forms of perspective-taking appear to occupy the "spontaneous" middle ground between the fast-yet-inflexible and flexible-yet-slow information-processing profiles of the implicit and explicit mindreading systems. In both cases, this combination of flexibility and efficiency seems to be achieved through the interaction between executive systems, long-term

[12] See Thompson (2014) for a detailed critique of this proposal.

memory, and motivational factors. This is not to say that the underlying processes in the two kinds of perspective-taking are really identical: both seem to make use of different cognitive strategies, and are suited to different types of problems. But neither are the two clearly dissociable, as the two-systems framework would suggest.

One obvious conclusion to be drawn from this fact is that there need not be any trade-off between speed and representational flexibility when it comes to our perspective-taking abilities. On its own, this conclusion may not be fatal to the two-systems account: perhaps the distinction between Level-1 and Level-2 perspective-taking does not map onto the implicit and explicit mindreading systems after all, but this framework may still capture important distinctions when it comes to other forms of mental-state attribution. However, the case of perspective-taking should also lead us to view the basic idea of a flexibility-efficiency trade-off in the domain of mindreading with suspicion. Not only does this notion of a trade-off not apply in the case of perception – the domain it was originally intended to explain – but now it has also fallen short in explaining the cognitive underpinnings of one of our core mindreading abilities. Why expect that it should suddenly apply elsewhere?

As a matter of fact, there is evidence that in addition to Level-1 perspective-taking, other forms of "implicit" mindreading also appear to be unencapsulated from background knowledge. For instance, the attribution of motor intentions[13] through motor simulation or "mirror neurons" is often suggested to be automatic and encapsulated from background knowledge. Most commonly, this process is said to involve the automatic mapping of the visual kinematics of an observed action onto the motor system. Our motor system then simulates the performance of that same action, which permits an inference to a guiding motor intention (Rizzolatti and Craighero 2004) by using our motor planning system in reverse (Jeannerod et al. 1995). According to this view, the only inputs to the mirror neuron system are the low-level visual properties of actions.

However, other research on the mirror neuron system is inconsistent with this picture. Monkeys' mirror neurons do not activate for mimicked actions, as when they observe an experimenter pretending to grasp a non-existent object (Gallese and Goldman 1998);

[13] Motor intentions are intentions to engage in a particular motor action, such as grasping or throwing. These are distinct from distal or future intentions (what I plan to do at some point in the future) and present intentions (what I plan to do now, framed at a level of abstraction that is independent of any particular motor plan) (Pacherie 2008; Spaulding 2015).

conversely, monkeys' mirror neurons do activate when they witness an occluded grasping action that has no low-level visual properties – but only if they know in advance that there is food behind the occluder (Umiltà et al. 2001). In humans, it's been found that background knowledge about whether or not an observed action is intentional, or whether it has been carried out by an intentional agent, affects the degree to which they are motor-primed to perform that same action (an effect of mirror neuron activity) (Liepelt and Brass 2010; Liepelt and Cramon 2008). In other words, the attribution of motor intentions, just like the attribution of Level-1 perspectives, does not seem to be fully automatic or informationally encapsulated. Rather, it is sensitive to background knowledge and abstract features of context. Several authors have taken these findings as evidence that the mirror neuron system actually reflects the effects of a top-down, information-rich form of action prediction, rather than a low-level mapping process (Gergely and Csibra 2008; Kilner and Frith 2007).

Further problems for the two-systems account of mindreading arise from studies of "implicit" false-belief[14] tracking (Schneider, Bayliss, et al. 2012). In these tasks, subjects in an eye-tracker passively observe videos of an agent hiding an object and then leaving a room. While the agent is absent, the location of the object is changed. When she returns, subjects look in anticipation towards the previous location of the hidden object (the one last seen by the agent), suggesting that they were tracking her false beliefs. When subjects were debriefed after the task, they showed no sign that they were consciously tracking the agent's belief, suggesting that any belief-tracking that occurred was unconscious and implicit. However, when subjects in the same task are given even a light working-memory task, the implicit belief-tracking effect vanishes (Schneider, Lam, et al. 2012). One way of interpreting this finding is to conclude that implicit belief-tracking involves working memory; however, given that the contents of working memory are usually conscious, and subjects reported no conscious belief-tracking, this seems unlikely. What's more plausible is that when subjects were engaged in the working memory task, they shifted too much attention away from the agent's perspective for encoding of belief-states to occur or be fixed in long-term memory. Thus, implicit belief-tracking does not seem to be genuinely automatic; rather, as Level-1 perspective-taking, it's likely that we possess a

[14] Proponents of the two-systems account would deny that these experiments provide evidence for "belief-tracking," since they hold that the implicit system does not represent "full-blown" propositional attitudes. Rather, they would describe these results as evidence of the tracking of "registrations," a quasi-mentalistic, implicit analogue of beliefs represented by the implicit system (Butterfill and Apperly 2013).

standing disposition to represent the beliefs of others, but only when doing so is either cognitively efficient or somehow goal-relevant.

These findings suggest that other forms of implicit mindreading may also be spontaneous and context-sensitive, rather than automatic and encapsulated. If so, then the entire two-systems framework may be in jeopardy. The principal theoretical motivation for the two-systems account was that fast, efficient, "implicit" processes are likely to be encapsulated, which in turn yields signature limits on their representational capabilities. Instead, we find that implicit mindreading processes are generally quite flexible, and well-integrated with long-term memory, executive systems, and goals. If this is right, then it's not obvious whether there really are any grounds for holding the implicit mindreading system exists.

If the implicit mindreading system is not present in adults, this also casts doubt on the developmental claims of the two-systems view. Part of the two-systems approach to development has been to propose that younger children's early theory-of-mind abilities (e.g. Onishi & Baillargeon, 2005) are products of the implicit mindreading system, and thus subject to "signature limits" on their representational abilities (Butterfill and Apperly 2013); in particular, children below the age of four should not be able to pass Level-2 perspective-taking tasks, since these require "full blown" propositional attitude attribution. Proponents of the two-systems account tested this prediction in two separate studies, and obtained seemingly positive results: infants' looking times did not reflect any Level-2 perspective-taking, and thus seemed subject to signature limits (Low and Watts 2013; Low et al. 2014). But as with other Level-2 perspective-taking tasks, these paradigms involved mental rotation, and thus potentially confound Level-2 perspective-taking with effortful uses of working memory (Carruthers 2015c).

When this mental-rotation objection is supplemented by the revelation that the "signature limits" interpretation is based on an erroneous, encapsulated conception of the implicit mindreading system, it becomes all the more clear that these results provide no support for a two-systems account of infant theory-of-mind abilities. If infant mindreading abilities are really subject to any signature limits on their representational capabilities, it is unlikely that these are due to a distinct, encapsulated mindreading system that persists into adulthood. These limitations are more likely the product of immature executive resources, motivational factors, or a lack of relevant experience. Collectively, these factors may create a kind of ersatz

encapsulation early in development, but this would dissipate as children's developing executive resources and increasing social experience provides them with a more flexible, integrated set of mindreading abilities.

7. Conclusion: Efficient, context-sensitive mindreaders

Beyond its implications for the two-systems account, this critique highlights some important features of our mature mindreading abilities. One is that several implicit forms of mindreading do not seem to be genuinely automatic; rather, we deploy these capacities selectively, in a context-sensitive, goal-dependent fashion (although we may also be generally motivated to engage in mentalizing when doing so is cognitively efficient). However, our context-sensitive, goal-dependent mindreading abilities can still be quite fast and efficient. This combination of speed and context-sensitivity seems to be due to the integration of domain-specific mindreading mechanisms with domain-general attentional processes and background knowledge. We also find that even complex, so-called "explicit" forms of mental-state attribution, such as Level-2 perspective-taking, can also be both fast and efficient, provided that we possess the right background knowledge and that we are appropriately motivated.

Another significant conclusion to draw from this discussion is that whether we spontaneously engage in very simple forms of mindreading, or very complex forms of mindreading, or no mindreading at all, seems to be a function of our motivations. Along with our background knowledge, our social attitudes seem to determine the amount of processing resources that go into representing the minds of others. Sometimes, we are highly motivated to represent the mental states of others accurately, and we make use of background knowledge in order to do so quickly and efficiently; at other times, we are less motivated, and as a result our mental state representations are much sparser, as we rely on general-purpose heuristics, such as computing line-of-sight. And, as we saw in many of the gaze-cueing studies, sometimes our background beliefs about the intentional status of an agent or its social group membership give us reason to ignore its perspective altogether. The depth of processing involved in a given mindreading task thus depends on our social goals.

Moreover, as we saw in the discussion of Elekes et al. (2016) and Surtees et al. (2016), the availability of relevant background knowledge enables the mindreading system with a way to circumvent slower, more effortful forms of reasoning. Notably, in these studies, the relevant background knowledge was not antecedently available to the participants when they first

engaged in the task. But when subjects were sufficiently motivated to do so, they were able to generate situation-specific, mentalistic schemas that enabled them to rapidly update their representation of their partner's mental states. In other words, one of the things that we seem to do during social interactions is create shortcuts that make the task of mindreading faster and more efficient – provided, that is, that we are motivated to do so.

Now, contrast this picture of mindreading with the one put forward by mindreading skeptics and endorsed by two-systems theorists (Apperly 2011; Bermudez 2003; Zawidzki 2013). On their view, genuine mental-state attribution consists in a holistic, unbounded form reasoning that parallels the structure of first-person decision-making. According to this picture, mindreaders must, when inferring the mental cause of an action, consider an indefinite range of potential belief-desire combinations. The computational demands of this kind of mental-state inference are surely immense. Clearly, as a theory of how we are able to seamlessly engage in complex forms of coordination or quickly infer intended speaker meanings, this model of mindreading is inadequate; rather, it seems to represent the mental-state attribution strategy of an ideal thinker, unhindered by the demands of computational complexity.

Not being ideal thinkers ourselves, we rarely – if ever – engage in this kind of mindreading. But, contrary to the mindreading skeptic, this does not mean that we rarely engage in mindreading at all. Nor does it mean that we rely on a module for quasi-mentalistic mindreading, as the two-systems theorists have proposed. Rather, we deploy a range of flexible mentalizing strategies to navigate the social environment, which we tailor to match our situational needs. Some of these strategies may indeed involve effortful, working-memory based forms of cognition. But we do not engage in these effortful reasoning strategies any more than is necessary. Instead, we supplement this kind of reasoning with a number of more efficient strategies. Sometimes, these involve simple, spatial heuristics, as with Level-1 perspective-taking. But we also use more effortful forms of reasoning to create mindreading shortcuts, in the form of mentalistic schemas that may be rapidly retrieved from memory in order to maintain up-to-date models of other people's mental states. And even these more efficient forms of mindreading are deployed in a selective, context-sensitive manner. In short, we economize our mindreading strategies so that they may best fit our needs. We only ever mindread as much as we have to.

Thus, skeptical doubts about the mindreading paradigm can be assuaged once we appreciate the context-sensitive, goal-dependent nature of mental-state attribution. It is a mistake to believe that everyday mindreading consists in a holistic, unbounded form of "central" reasoning. It is also a mistake to argue that if we rarely engage in this idealized form of mindreading, then we must not mindread very much at all. The two-systems view attempted to carve out a middle ground between these two extremes, but it erred in its concession to the skeptic that "full-blown" mindreading must cognitively effortful. With the case of spontaneous perspective-taking, I've shown that our mindreading abilities are much more flexible, efficient and context-sensitive than either the two-systems theorists and the skeptics had thought possible.

References:

Apperly, I. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind."* Psychology Press.

Apperly, I. (2013). Can theory of mind grow up? Mindreading in adults, and its implications for the development and neuroscience of mindreading. In S. Baron-Cohen, H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (3rd ed., pp. 72–92). Oxford: Oxford University Press.

Apperly, I., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970.

Baillargeon, R., Scott R.M., & He, Z. (2010). False-belief Understanding in Infants. *Trends in Cognitive Sciences, 14*(3), 110–118.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103*(2), 449–454.

Bermudez, J. L. (2003). The Domain of Folk Psychology. *Royal Institute of Philosophy Supplement*, 25–48.

Bratman, M. (1992). Shared Cooperative Activity. *Philosophical Review, 101*(2), 327–341.

Bräuer, J., Call, J., & Tomasello, M. (2004). Visual perspective taking in dogs (Canis familiaris) in the presence of barriers. *Applied Animal Behaviour Science, 88*(3-4), 299–317.

Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature communications, 7*, 10506.

Butterfill, S., & Apperly, I. (2013). How to Construct a Minimal Theory of Mind. *Mind and Language, 28*(5), 606–637.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in cognitive sciences, 12*(5), 187–92.

Carruthers, P. (2013). Mindreading in Infancy. *Mind & Language, 28*(2), 141–172.

Carruthers, P. (2015a). *The centered mind: what the science of working memory shows us about the nature of human thought.* Oxford University Press.

Carruthers, P. (2015b). Mindreading in adults: evaluating two-systems views. *Synthese, 192*, 1–16.

Carruthers, P. (2015c). Two Systems for Mindreading? *Review of Philosophy and Psychology, 6.*

Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2014). Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral cortex, 24*(11), 2899–907.

Chen, Y., & Zhao, Y. (2015). Intergroup threat gates social attention in humans.

Christensen, W., & Michael, J. (2015). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology, 40*(A), 48–64.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306–24.

Dalmaso, M., Pavan, G., Castelli, L., & Galfano, G. (2012). Social status gates social attention in humans. *Biology Letters*, *8*(3), 450–452.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the cognitive sciences*, *6*(4), 485–507.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93–103.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2009). Gaze Following in Newborns. *Infancy*, *5*(1), 39–60.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Fiebich, A., & Coltheart, M. (2015). Various Ways to Understand Other Minds: Towards a Pluralistic Approach to the Explanation of Social Understanding. *Mind and Language*, *30*(3), 235–258.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.

Fodor, J. (1992). A theory of the child's theory of mind. *Cognition*, *44*(3), 283–296.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*(2), 535–543.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the mind-reading. *Trends in cognitive sciences*, *2*(12), 493–501.

Gergely, G., & Csibra, G. (2008). Action mirroring and action understanding: an alternative account. *Sensorymotor Foundations of Higher Cognition. Attention and Performance XXII*, 435–459.

Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Grice, H. P. (1991). *Studies in the Way of Words*. Harvard University Press.

Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *9*(2), 131–43.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*, *9*(2), 131–134.

Hutto, D. D. (2012). *Folk psychological narratives: The sociocultural basis of understanding reasons*. MIT Press.

Hyun, J.-S., & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation.

Psychonomic Bulletin & Review, *14*(1), 154–158.

Jakobsen, K. V., Frick, J. E., & Simpson, E. A. (2013). Look Here! The Development of Attentional Orienting to Symbolic Cues. *Journal of Cognition and Development*, *14*(2), 229–249.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objets: the cortical mechanisms of visuomotor transformation. *Trends in Neurosciences*, *18*(7), 314–320.

Kilner, J. M., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processes*, *8*(3), 159–166.

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and Cognition*, *55*(2), 269–271.

Knudsen, E. I. (2011). Control from below: the role of a midbrain network in spatial attention. *The European journal of neuroscience*, *33*(11), 1961–72.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in cognitive sciences*, *8*(12), 528–33.

Lewis, D. (1969). *Convention: A philosophical study*. John Wiley & Sons.

Liepelt, R., & Brass, M. (2010). Top-Down Modulation of Motor Priming by Belief About Animacy, *57*(3), 221–227.

Liepelt, R., & Cramon, D. Y. Von. (2008). What Is Matched in Direct Matching ? Intention Attribution Modulates Motor Priming, *34*(3), 578–591.

Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: limits on preschoolers' efficient ability to track perspective. *Child development*, *85*(4), 1519–34.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: state of the art. *The British journal of developmental psychology*, *30*(Pt 1), 1–13.

Low, J., & Watts, J. (2013). Attributing false-beliefs about object identity is a signature blindspot in humans' efficient mindreading system. *Psychological Science*, *24*(3), 305–311.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental science*, *12*(1), 142–9.

Marotta, A., Lupiáñez, J., Martella, D., & Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: two qualitatively different modes of attentional selection. *Journal of experimental psychology. Human perception and performance*, *38*(2), 326–35.

Masangkay, Z. S., McCluskey, K. a, McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child development*, *45*(2), 357–366.

Michael, J., & D'Ausilio, A. (2015). Domain-specific and domain-general processes in social perception – A complementary approach. *Consciousness and Cognition*, *36*, 434–437.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & psychophysics*, *68*(2), 327–337.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in cognitive sciences*, *11*(4), 143–52.

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*(3), 603–613.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

Morton, A. (1996). Folk Psychology is not a Predictive. *Mind*, *105*(417), 119–137.

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Clarendon Press/Oxford University Press.

Ogilvie, R., & Carruthers, P. (2016). The case against encapsulation. *Review of Philosophy and Psychology*, *7*.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–8.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179–217.

Pavan, G., Dalmaso, M., Galfano, G., & Castelli, L. (2011). Racial group membership is associated to gaze-mediated orienting in Italy. *PLoS ONE*, *6*(10).

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(03), 341–365.

Qureshi, A. W., Apperly, I., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–6.

Rakoczy, H. (2015). In defense of a developmental dogma: children acquire propositional attitude folk psychology around age 4. *Synthese*.

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*(3), 507–513.

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition*, *94*(3).

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience*, *27*, 169–92.

Samson, D., Apperly, I., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: implicit mentalizing or domain-general processing? *Journal of experimental psychology. Human perception and performance*, *40*(3), 929–37.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *141*(3), 433–438.

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological science*, *23*(8), 842–7.

Scholl, B. J., & Leslie, A. M. (1999). Modularity , Development and " Theory of Mind ." *Mind & Language*, *14*(1), 131–153.

Slessor, G., Laird, G., Phillips, L. H., Bull, R., & Filippou, D. (2010). Age-related differences in gaze following: does the age of the face matter? *The journals of gerontology. Series B, Psychological sciences and social sciences*, *65*(5), 536–41.

Spaulding, S. (2015). On Direct Social Perception. *Consciousness and Cognition*, *36*, 472–482.

Surtees, A., Apperly, I., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426–438.

Surtees, A., Apperly, I., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in human neuroscience*, *7*(November), 698.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52.

Surtees, A., Butterfill, S., & Apperly, I. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *The British journal of developmental psychology*, *30*(Pt 1), 75–86.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105.

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, perception & psychophysics*, *72*(3), 695–705.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(11), 4567–70.

Thompson, J. R. (2014). Signature Limits in Mindreading Systems. *Cognitive Science*, *38*(7), 1432–1455.

Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *Monist*, *59*(2), 204–217.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and brain sciences*, *28*(5), 675–91; discussion 691–735.

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I Know What You Are Doing. *Neuron*, *31*(1), 155–165.

Wellman, H. M. (2014). *Making Minds: How Theory of Mind Develops*. Oxford: Oxford

University Press.

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PloS one*, *7*(9), e45391.

Wilson, D., & Sperber, D. (2012). *Meaning and Relevance*. Cambridge University Press.

Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in psychology*, *5*, 674.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(20), 8235–40.

Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. MIT Press.