How to Construct a Minimal Theory of Mind

Stephen A. Butterfill & Ian A. Apperly
<s.butterfill@warwick.ac.uk>

April 6, 2012

Abstract

What could someone represent that would enable her to track, at least within limits, others' perceptions, knowledge states and beliefs including false beliefs? An obvious possibility is that she might represent these very attitudes as such. It is sometimes tacitly or explicitly assumed that this is the only possible answer. However we argue that several recent discoveries in developmental, cognitive, and comparative psychology indicate the need for other, less obvious possibilities. Our aim is to meet this need by describing the construction of a minimal theory of mind. Minimal theory of mind is rich enough to explain systematic success on tasks held to be acid tests for theory of mind cognition including many false belief tasks. Yet minimal theory of mind does not require representing propositional attitudes, or any other kind of representation, as such. Minimal theory of mind may be what enables those with limited cognitive resources or little conceptual sophistication, such as infants, chimpanzees, scrub-jays and human adults under load, to track others' perceptions, knowledge states and beliefs.

Keywords: Theory of Mind, False Belief, belief, perception, development, comparative

1. Introduction

What could someone represent that would enable her to track, at least within limits, others' perceptions, knowledge states and beliefs including false beliefs? One answer is obvious: she might track these things by virtue of representing them as such, that is, by representing perceptions, beliefs, and other

propositional attitudes as such. Our aim in what follows is to identify another, less obvious answer. There is a form of cognition—minimal theory of mind—which does not involve representing propositional attitudes as such but does involve representing simpler, relational mental states which could, within limits, enable one to track propositional attitudes such as beliefs. Minimal theory of mind is rich enough to enable systematic success on tasks held to be acid tests for theory of mind cognition including many false belief tasks. As we will explain, this has consequences for interpreting a range of findings concerning infants', adults' and nonhumans' performances on theory of mind tasks. It may help us to understand what enables those with limited cognitive resources or little conceptual sophistication, such as infants, chimpanzees, scrub-jays and human adults under load, to track, within limits, facts about perceptions and beliefs.

In this section we defend explain our question; in the next sections we introduce the findings which motivate facing it before starting to answer it in the fourth section.

Some may find our question initially incomprehensible. Could abilities to track false beliefs (say) really involve anything other than representing false beliefs? To see the possibility of a positive answer it may help to consider a non-mental analogy. What could someone represent that would enable her to track, at least within limits, the toxicity of potential food items? Here the most straightforward answer (she could represent their toxicity) is clearly not the only one. After all, someone might track toxicity by representing odours or by representing visual features associated with putrefaction, say. Suppose Sinéad has no conception of toxins but represents the odours of food items and treats those with foul odours as dangerous to eat, so that she would not normally offer them to friends or family nor conceal them from competitors. This brings nutritional and competitive benefits obtaining which depends on facts about toxicity. If Sinéad tends to behave in this way because of these benefits, representing odours enables her to track, in a limited but useful range of cases, toxicity. Our question, put very roughly, is whether belief has something like an odour.

To make the question more precise it is useful to distinguish theory of mind abilities from theory of mind cognition. A *theory of mind ability* is an ability that exists in part because exercising it brings benefits obtaining which depends on exploiting or influencing facts about others' mental states. To illustrate, suppose that Hannah is able to discern whether another's eyes are in view, that Hannah exercises this ability to escape detection while stealing from others, that Hannah's ability exists in part because it benefits her in this way, and that Hannah's escaping detection depends on exploiting a fact about other's mental states (namely that they usually cannot see Hannah's acts of theft when Hannah doesn't have their eyes in view). Then Hannah has a theory of mind ability. (This is not supposed to be a plausible, real-

world example but only to illustrate what the definition requires.) An ability to *track* perceptions or beliefs (say) is a theory of mind ability which involves exploiting or influencing facts about these states. By contrast, *theory of mind cognition* involves representing mental states or processes as such. And *full-blown* theory of mind cognition involves representing propositional attitudes such as beliefs, desires and intentions to construct reason-giving, causal explanations of action. The distinction between theory of mind abilities and theory of mind cognition matters because the facts about other minds which theory of mind abilities exploit are not necessarily the facts which are represented in theory of mind cognition. To return to the illustration, Hannah is able, within limits, to exploit facts about what others perceive without representing perceptions as such. She has a theory of mind ability while possibly lacking any theory of mind cognition.

It should be uncontroversial that some theory of mind abilities do not necessarily involve any theory of mind cognition at all. Our question concerns abilities to track what others perceive and believe, including their false beliefs; these have been central in psychological research. Can anything less than full-blown theory of mind cognition explain systematic success on a range of false belief tasks? We do not aim to argue that someone could track beliefs, true and false, without any theory of mind cognition at all. Our concern is rather with the construction of a minimal form of theory of mind cognition. As we shall explain, minimal theory of mind does involve representing belief-like states, but it does not involve representing beliefs or other propositional attitudes as such.

The notion that some abilities to track perceptions, knowledge states or beliefs involve only theory of mind cognition which does not involve representing perceptions or beliefs as such is not entirely novel. To mention only those we draw most directly on, Gomez (2007, p. 730) has emphasized primitive intentional relations to objects established by gaze, O'Neil and Doherty have separately discussed a notion of engagement with objects (Doherty 2006; O'Neill 1996), Call and Tomasello (2005, p. 58) have suggested that chimpanzees track the 'likely target' of others' visual access and understand something about its effects on behaviour, and Whiten (1994; 1996) uses the notion of an 'intervening variable' to explain primitive theory of mind notions. These are illuminating ideas and what follows can be seen as an elaboration and partial synthesis of them. The result—minimal theory of mind—is unlike its precursors in that it is rich enough to explain systematic success on a range of false belief tasks. Our approach is novel in this and others respects to which we return (in the Conclusion) after presenting the substance of our account.

2. Motivation

Our question is theoretical: it concerns not what anyone does represent but what someone could represent that would enable her, at least within limits, to track perceptions, knowledge states and beliefs. The motivation for facing up to this question is, of course, partly empirical.

Consider ordinary adult humans. Since they can represent beliefs and other propositional attitudes as such, it is natural to assume that such representations underpin their abilities to track perceptions and beliefs. But is this natural assumption correct?

To see that it might not be, consider a further question. Is tracking others' perceptions and beliefs automatic? Roughly speaking, a process is *automatic* if whether it occurs is to a significant degree independent of its relevance to the particulars of the subject's motives and aims. (Note that a process may occur spontaneously without thereby being automatic.) Some evidence suggests that, for ordinary adult humans, belief tracking is automatic. For example, Kovács et al. (2010) asked adults to identify the location of a ball. They found that adults were significantly slower to identify the ball's location when an onlooker had a false belief about the location of the ball, even though the onlooker's belief was not relevant to the task at all. Relatedly, Samson et al. (2010) provide evidence that identifying what another perceives is automatic; this finding is indirectly supported by evidence that tracking others' perceptions is not disrupted by a secondary executive task (Qureshi et al. 2010a). Taken together, these findings suggest that, at least in adults, tracking others' perceptions and beliefs is sometimes automatic.

But there is also a body of evidence supporting a different conclusion. Back & Apperly (2010) found that subjects are significantly slower to answer an unexpected question about another's belief when that belief is false compared to when it is true (see also Apperly et al. 2006). This suggests that, at least in adults, belief tracking is not automatic. There is also evidence that, even in relatively simple situations, using facts about others' beliefs is not automatic (Keysar et al. 2003; Apperly et al. 2010). The case for nonautomaticity is indirectly supported by evidence that tracking perceptions and beliefs—and even merely holding in mind what another believes, where no inference is required—involves a measurable processing cost (Apperly et al. 2008, 2010), consumes attention and working memory in fully competent adults (Apperly et al. 2009; Lin et al. 2010; McKinnon & Moscovitch 2007 experiments 4-5), may require inhibition (Bull et al. 2008) and makes demands on executive function (Apperly et al. 2004; Samson et al. 2005). These findings, taken together, suggest that tracking others' perceptions and beliefs is sometimes not automatic.

The question was whether, in adult humans, tracking perception and belief is automatic. If we assume, further, that either all such processes are

automatic or else none are, then the evidence creates a conflict. This conflict cannot easily be explained away by appeal to simple methodological factors or extraneous task demands. For instance, it may be tempting to suppose that the conflict can be explained by distinguishing between linguistic and nonlinguistic tasks. But belief ascription may fail to be automatic even on some nonlinguistic tasks (e.g. Apperly et al. 2004), and we know of no reason to assume that belief ascription could not be automatic on some linguistic theory of mind tasks (such as those where spontaneous tracking is already established, e.g. Ferguson & Breheny (2012)).

If the conflict is not a methodological artefact, how should we interpret the evidence? Perhaps it should be taken at face value. This means we must reject the assumption that tracking others' perceptions and beliefs is either always automatic or else always nonautomatic. In other cases, such as number and causation, it is already quite widely accepted that, in adult humans, some abilities to track these things are automatic whereas others are not.[1] The evidence suggests that the same may be true for perception and belief. In adult humans, some theory of mind abilities involve automatic processes whereas others depend on nonautomatic processes.

A closely related view has already been elaborated and defended in more detail by Apperly & Butterfill (2009), although their argument complements ours by drawing primarily on developmental and comparative research. According to them, adults may enjoy efficient but inflexible forms of theory of mind cognition in addition to the full-blown form which involves representing beliefs and other propositional attitudes as such. While aspects of this conjecture have already been tested (Samson et al. 2010; Schneider et al. 2011; Surtees et al. 2011), it raises two complementary questions (as Apperly & Butterfill themselves note).

First, why isn't tracking belief and perception always automatic? Consider what is involved in representing beliefs and other propositional attitudes. 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 (Davidson 1980, 1990). If anything should consume working memory and other scarce cognitive resources, it is surely representing states with this combination of properties. So even without knowing in any detail how theory of mind cognition is implemented, it is plausible that some feature, or combination of features, of the propositional attitudes makes full-blown theory of mind cognition demanding.[2] A possible

[1] On number: Trick & Pylyshyn (1994); on causation: Michotte (1963), Scholl & Nakayama (2004).

[2] Several hypotheses about which feature of the propositional attitudes explains why full-blown theory of mind cognition is cognitively and conceptually demanding have been defended (e.g. Birch & Bloom 2007; Doherty & Perner 1998; Leslie et al. 2005; Lewis 1994;

explanation, then, is this. Tracking perception or belief is not always automatic because it sometimes involves representing propositional attitudes as such, which typically or always places demands on working memory, attention and executive function that are incompatible with automaticity.

Second, how could tracking perceptions or beliefs ever be automatic? If we assumed that such tracking always involved propositional attitudes as such, this question would present a puzzle. For, as we saw, representing propositional attitudes as such generally places demands on working memory, attention and executive function that are incompatible with automaticity. In some cases these demands might be overcome through automatization in something like the way that initially effortful numerical operations can through practice become automatic.[3] However, almost nothing is known about to what extent, if any, automatization occurs in theory of mind. And in any case automatization can only explain the automaticity of routine inferences. So it is possible that automatization, although perhaps important, does not fully explain the automaticity of some of adult humans' perception- and belief-tracking abilities. A full explanation may depend on showing that tracking perceptions and beliefs can be done without representing beliefs or other propositional attitudes as such.

This is a source motivation for our question about what someone could represent that would enable her to track perceptions and beliefs. The existence of both automatic and nonautomatic tracking of perceptions and beliefs in human adults suggests (without decisively showing, of course), contrary to a natural assumption mentioned above, that not all of their abilities to track perceptions and beliefs involve representing propositional attitudes as such.

3. More motivation

Further motivation for our question comes from evidence for theory of mind abilities in young children and infants. Children in their second year use pointing to provide information to others (Liszkowski et al. 2006) in ways that reflect a partner's ignorance or knowledge (Liszkowski et al. 2008), as well as providing more information to ignorant than knowledgeable partners when making requests (O'Neill 1996). One-year-old children also predict actions of agents with false beliefs about the locations of objects (Onishi & Baillargeon 2005; Southgate et al. 2007) and choose different ways of in-

Perner 1991; Perner et al. 2002; Russell et al. 1991; Sabbagh 2006). More than one feature may contribute. We are agnostic about which feature or features are to blame.

[3] On the automatization of simple sums, see LeFevre et al. (1988). For the suggestion that something similar might happen concerning mental states, see Suddendorf & Whiten (2003).

teracting with others depending on whether their beliefs are true or false (Buttelmann et al. 2009; Knudsen & Liszkowski 2012; Southgate et al. 2010). And in much the way that irrelevant facts about the contents of others' beliefs modulate adult subjects' response times, such facts also affect how long 7-month-old infants look at some stimuli (Kovács et al. 2010).

What do these infants and young children represent that enables them, within limits, to track others' perceptions, knowledge states and beliefs? The most straightforward answer would be to suppose that they represent perceptions, knowledge states and beliefs as such (e.g. Leslie 2005; Song et al. 2008). But this answer faces several objections. A body of evidence suggests that representing beliefs requires conceptual sophistication, for it has a protracted developmental course stretching over several years (Wellman et al. 2001; Wimmer & Perner 1983) and its acquisition is tied to the development of executive function (Perner & Lang 1999; Sabbagh 2006) and language (Astington & Baird 2005). Infants and young children are deficient in these. Development of reasoning about beliefs in humans may also be facilitated by explicit training (Slaughter & Gopnik 1996) and environmental factors such as siblings (Clements et al. 2000; Hughes & Leekam 2004). This is evidence that representations of belief in humans typically emerge from extensive participation in social interactions (as Hughes et al. 2006 suggest). If any of this is right, we must reject the hypothesis that infants are representing beliefs or other propositional attitudes as such.

In principle an alternative would be to suppose that infants' and young children's abilities to track perceptions and beliefs do not involve any theory of mind cognition at all but are instead based on representations of nonintentional behaviour only. It is arguably possible in principle to explain some belief-tracking abilities by appeal to hypothetical behaviour reading capacities (Perner 1990; Perner & Ruffman 2005; Ruffman & Perner 2005). But there are several objections to the claim that the full range of even infants' abilities to track perceptions and beliefs could be explained in this way (Song et al. 2008; Apperly & Butterfill 2009). And what is is currently known about humans' actual behaviour reading capacities suggests that they are unlikely to explain systematic success on false belief tasks.[4]

Here, then, is a second source of motivation for our question about what someone could represent that would enable her, within limits, to track perceptions and beliefs. As we have seen, there are significant if not decisive objections to the two best developed conjectures about infant theory of mind abilities, the conjecture that these involve representing perceptions, knowledge states and beliefs as such and the conjecture that these involve representing nonintentional behaviour only. These objections, while not decisive,

[4] Key studies include Newtson & Engquist (1976), Byrne (1999), Baldwin & Baird (2001), Saylor et al. (2007) and Baldwin et al. (2008).

justify exploring alternatives.

Theory of mind abilities are not only found in humans. For instance, scrub-jays can selectively re-cache their food in ways that deprive competitors of knowledge of its location (Clayton et al. 2007), and chimpanzees can both select routes to approach food which conceal them from a competitor's view (Hare et al. 2006) and also retrieve food using strategies that optimize their return given what a dominant competitor has seen (Hare et al. 2001). There is debate about the cognitive underpinnings of these abilities. Some researchers argue that they may involve representations of nonintentional behaviour only,[5] others that they involve representations of perceptions, knowledge states and other propositional attitudes.[6] If these arguments are not yet decisive and if the available evidence does not already tightly constrain the space of admissible hypotheses, then the construction of a minimal theory of mind may also be relevant to these debates. For the hypothesis that minimal theory of mind explains particular subjects' abilities to track perceptions or beliefs can be empirically distinguished from hypotheses about representations of nonintentional behaviour only and also from hypotheses about representations of perceptions, knowledge states and beliefs (as we explain in Sections 5 and 6 below).

4. Minimal theory of mind

In this section we begin with someone, call her Lucky, capable of representing nonintentional behaviour only and ask what more is needed for minimal theory of mind cognition. We describe Lucky's progress with a series of principles. The principles are constructed in such a way that it would be coherent to suppose that Lucky has the abilities codified by the first n principles only. They are not intended to represent a developmental or evolutionary progression. The principles can also be extended to explain more sophisticated theory of mind abilities than those considered here. We restrict ourselves to these five principles because they are sufficient to explain success on some false belief tasks.[7]

---

[5] For example, on chimpanzees: Povinelli & Vonk (2004); Vonk & Povinelli (2006, pp. 364-5); and on scrub-jays and chimpanzees Penn & Povinelli (2007).

[6] For example, on chimpanzees: Tomasello & Call (2005); Call & Tomasello (2008); and on scrub-jays: Emery & Clayton (2007, p. 73).

[7] In standard false belief tasks, '[t]he subject is aware that he/she and another person [call him Maxi] witness a certain state of affairs x. Then, in the absence of the other person the subject witnesses an unexpected change in the state of affairs from x to y' (Wimmer & Perner 1983, p. 106). The test concerns whether the subject realises that Maxi will falsely believe x to obtain. In many cases the states of affairs, x and y, differ only with respect to the location of an object (e.g. Onishi & Baillargeon 2005; Southgate et al. 2007; Träuble et al. 2010). As we go on to discuss, our proposal for a minimal theory of mind could

We aim to provide the core elements of a computational theory in Marr's sense (1982, pp. 15-29) where our computational theory, unlike the standard full-blown theory of mind which hinges on beliefs, desires and other propositional attitudes, is one that could be realised in a cognitively efficient manner without requiring conceptual sophistication.[8] There are multiple ways in which this computational theory might be implemented. We shall not discuss how the theory might be implemented here other than to note that it seems unlikely that the principles formulated below are represented explicitly. It is valuable to articulate the computational theory in some detail before formulating and testing conjectures about implementation.

4.1. First principle

The first principle concerns goal-directed action. The term 'goal-directed action' can be used to mean several things. One is intentional action. To represent intentional actions as such you also have to represent intentions or propositional attitudes such as beliefs and desires (Davidson 1999a). This notion is therefore no use for constructing a minimal theory of mind—our aim is to explain how Lucky could track perceptions and beliefs without representing beliefs or other propositional attitudes as such. Instead we need a more basic notion of goal-directed action.

A suitable notion of goal-directed action can be chracterised in terms of functions (Taylor 1964; Dretske 1988). The units of goal-directed action are events comprising mere bodily movements. We stipulate that for an outcome, *g*, to be the goal of some bodily movements is for these bodily movements to occur in order to bring about *g*; that is, *g* is the function of this collection. Here 'function' should be understood teleologically. On the simplest teleological construal of function, for an action to have the function of bringing about *g* would be for actions of this type to have brought about *g* in the past and for this action to occur in part because of this fact (see further Godfrey-Smith 1996; Millikan 1984; Neander 1991; Price 2001; Wright 1976). Lucky needs some ability to track the functions of things (in this special sense of 'function') so that she can link some bodily movements to the goals to which they are directed.[9]

---

be extended to cover a range of other cases; but importantly there are also false belief tasks success on which cannot be explained by minimal theory of mind cognition (see Section 5 on page 18).

[8] Computational theories in Marr's sense are not necessarily implemented by computational processes.

[9] Note that the requirement is not that Lucky understands the theoretical account of functions, only that she can distinguish between things which have different functions in this theoretical sense of 'function'. A wide variety of research supports the claim that young children, non-human primates and corvids track the functions of things (includ-

This is not supposed to be a fully adequate account of goal-directed action. For our purposes what matters is not whether the account correctly identifies what goal-directed action is, but rather that it characterises what someone who has only a minimal grasp of goal-directed action might understand. For comparison, consider an individual whose understanding of kinship could be characterised by an incorrect theory of social relations. In practice it may not matter that the individual fails to fully understand kinship providing that she can reliably identify who is whose kin in her everyday life. Similarly, Lucky does not need to fully understand goal-directed action in order to be able to pick out, in a limited but useful range of cases, which bodily movements are directed to which outcomes.

Note that representing goal-directed action as we have characterised it does not require representing representations. It only requires representing outcomes as functions of bodily movements. (The term 'goal' is sometimes used, perhaps improperly, to refer to intentions or other representations; but as we use the term, a goal is simply an outcome to which an action is directed.)

The first principle, then, is that bodily movements form units which are directed to goals. This first principle is sufficient to explain some cases of imitative learning, which can be defined as attempting to reproduce the actions necessary to achieve a goal (Tomasello 1999).

4.2. Second principle

Before describing the second principle we need to introduce two concepts. An agent's *field* at any given time is a set of objects. Whether an object falls within the agent's field is determined by spatial and physical constraints such as proximity and lighting. The agent's orientation and posture will also play a role in determining which objects fall into an agent's field, as will eye direction in some species. To fall within an agent's field, there must be no opaque barriers between the agent and the object, unless the object was recently in motion and not behind a barrier. These constraints ensure that objects which fall into an agent's field are approximately those the agent can perceive.[10]

Let us say that an agent is *encountering* an object if it is in her field. The notion of encountering defines a relation between an agent and an object. Within limits, this notion of encountering can do some of the work that the

ing Rakoczy et al. 2005, Casler & Kelemen 2007, Csibra & Gergely 2007, Kelemen 1999, German & Defeyter 2000 and Emery & Clayton 2004).

[10] A variety of research in spatial and motor cognition suggests that adult humans (and perhaps others) not only compute other agents' fields but also spontaneously locate objects within the spatial perspectives of other agents (e.g. Carlson-Radvansky & Irwin 1994; Frischen et al. 2009).

concept of perception does. Encountering an object is like perceiving one to the extent that both notions involve a relation between agents and objects, both notions have approximately the same extension (someone perceives an object just if she encounters it), and both notions are bound up with action, as we shall explain.

With these concepts in place, we can state the second principle: one cannot goal-directedly act on an object unless one has encountered it. More carefully, if an outcome involves a particular object and the agent has not encountered that object, then that outcome cannot be a goal of her actions. As with the other principles, this is plainly not a fact. What matters is just that, in a limited but useful range of cases, the principles collectively enable lucky to track perceptions and goal-directed actions.

The second principle has many applications. Someone who is aware of this principle can be motivated to prevent others from encountering her food even when they are not in a position to steal it immediately. Take scrub-jays. When choosing where to cache food in the presence of a competitor they prefer far to near, darker to lighter, and occluded to in-view locations (Clayton & Emery 2007; Dally et al. 2004). These scrub-jays may be trying to hinder future thefts, for these behaviours are not found when caching non-food items (Bugnyar et al. 2007) or when caching in the presence of a partner (Clayton et al. 2007; Emery & Clayton 2007, p. 514). Clayton and Emery note that '[s]uch skills suggest visual perspective taking—computing what another can or cannot see' (Clayton & Emery 2007). That is to say, they ascribed to scrub jays the concept of seeing. Another possibility is that scrub-jays compute encountering rather than seeing. Perhaps scrub-jays take having encountered food to be a condition for performing goal-directed actions targeting that food. If so they may be trying to minimize the chance that others will encounter their food in order to prevent future theft. Of course we are not claiming that this is the actual explanation of these findings. Our question is not about what scrub-jays actually represent but about what someone could represent that would enable her to track perceptions. Our claim is that the ability to track perceptions in the ways scrub-jays do could involve representing encounterings only.

For another application of the second principle, consider Hare, Call and Tomasello's finding that chimpanzees reliably adopt strategies which are appropriate given what dominant competitors know about the locations of food (Hare et al. 2001). In their 'uninformed' condition, a subordinate chimpanzee observed a food item being hidden while a dominant competitor's view was blocked (see Figure 1 on the next page). In this condition subordinates chose to approach the food significantly more often than in a control condition where the dominant competitor saw the food being hidden. This indicates that the subordinate chimpanzees were at least indirectly sensitive to facts about what the dominants had perceived. Several explanations of this find-

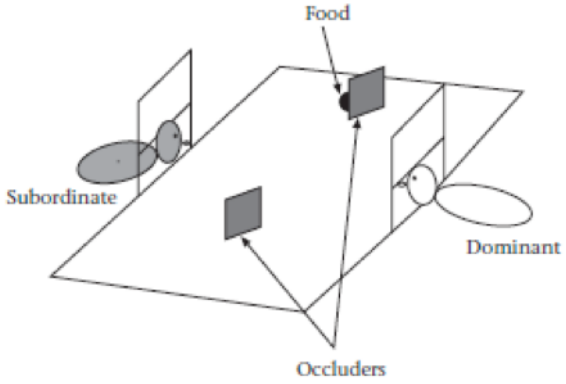

Figure 1: A subordinate observes as food is placed. The subordinate can also see the dominant. There are three conditions: control—the dominant sees food being placed; 'uninformed'—the dominant's view is blocked while the food is placed; and 'misinformed'—the dominant sees the food being placed then has their view blocked while it is moved. (Source: Hare et al. 2001, pp. 142, fig. 1)

ing have already been suggested (Call & Tomasello 2008; Povinelli & Vonk 2003; Suddendorf & Whiten 2003). A further possibility is that subordinate chimpanzees are aware that the dominant chimpanzee has not encountered the food and take encountering the food to be a condition for the dominant to act with the goal of recovering it. That would enable them to predict that the subordinate will not be able to retrieve this food in the misinformed condition.

In short, abilities to track others' perceptions may depend on representing perceptions as such. But another way to track perceptions would be to represent encounterings and to suppose (as the second principle states) that goal-directed actions involving an object are only possible when the agent has encountered that object.

So far we have been taking for granted that representing encounterings is different from representing perceptions as such. Why accept this claim? It is striking that philosophers have quite different views about what seeing involves—there are debates about whether it is possible to see things without seeing that something is the case (e.g. Dretske 1969), about whether and in what sense seeing is a representational notion (Martin 2002; Tye 2002), and about whether vision is intrinsically a source of knowledge (Cassam 2008), for example. There is at least as much uncertainty concerning what might be involved in representing perceptions as such. So it is not possible for us to say with certainty what it is to represent perceptions as such. Even so, we can be sure that there are some differences between encountering and per-

ception. If perceptions are representations, then representing perceptions as such plausibly involves representing representations. Since encounterings are relations not representations (by definition), representing encounterings will differ from representing perceptions in that only the latter involves representing representations. But, as mentioned, not everyone agrees that perceptions are representations. We should therefore consider a bundle of features. Perception constitutively involves appearances, modalities or the possibility of illusion or is constitutively linked to reasons, knowledge or informational states. It is plausible that representing perceptions as such involves understanding something about some of these things. Since encountering does not constitutively involve any of these (by definition), representing encounterings may differ from representing perceptions in that it does not require any kind of sensitivity to constitutive links with reasons, knowledge or informational states and nor does it require understanding anything about appearances, modalities or the possibility of illusion.

4.3. Third principle

At this point we switch our attention from conditions on the *occurrence* of goal-directed actions to conditions on their *success*. Such conditions are more stringent. Where a goal specifies a particular object, it's often not enough to have encountered it. To succeed you must often also have encountered it *in its current location* and to have done so *on your most recent encounter*. (There are exceptions, of course.)

We can capture this condition and more by introducing a new notion, *registration*. Registration is a relation between an individual, an object and a location which will be implicitly defined by principles linking it to encountering and action. By design, applying all of these principles would sometimes lead to inconsistencies. Where it would be impossible to apply all the principles consistently, consistency is to be achieved by having principles mentioned later trump principles mentioned earlier. The first principle defining registration is that an individual registers an object at a location if and only if she most recently encountered it at that location. As is already clear from this principle, registration is like belief in that it has a correctness condition which may not obtain: a registration is correct when the object is in the location. Since representing registration brings no benefit for Lucky independently of its connection to action, this principle cannot stand alone in our sequence; it is only the first half of the third principle.

The second half of the third principle states that correct registration is a condition of successful action. More precisely, in order to successfully perform a goal-directed action with a goal that specifies a particular object, the

agent must correctly register that object.[11] This principle can be applied in two directions. In one direction, it licenses Lucky to predict that a competitor who does not have a correct registration of an object will not be successful in performing actions whose goals specify that object. In the other direction, it allows Lucky, on the basis of observing a successful goal-directed action, to infer that the agent has correctly registered the location of an object.[12] So the principle not only extends Lucky's ability to predict actions but also her ability to detect what someone registers.

The correctness of someone's registrations can be manipulated in their absence by moving or destroying objects they have registered. So with theory of mind cognition partially characterized by the third principle, Lucky can intentionally prevent others from stealing a food item they have already encountered simply by moving it in their absence.

For an application of this principle, consider Hare, Call and Tomasello's (2001) experiment again. In a further condition, the 'misinformed' condition, a subordinate observer watched as a dominant competitor saw food being hidden. The subordinate continued to watch as the competitor's view was blocked and the food moved. In this case the competitor has encountered the food but does not correctly register it. Subordinate observers went for the food more often in this condition than in a control condition where the dominant saw the food being moved. This cannot be explained in terms of the second principle. That principle involved taking encountering an object to be a condition on acting on it. This condition is met: the competitor *has* encountered the food. To explain why the subordinate observer goes for the food that has been moved, we need to appeal to the third principle—to correct registration as a condition on success. It is possible that the subordinate observer realized that the dominant competitor last encountered the food in a location other than its current location. Suppose the observer also understood that correct registration is a condition on successful goal-directed action. Then the observer could predict that the competitor would not succeed in retrieving the food. This could explain why subordinate observers more often approach the food in the 'misinformed' condition than in the control condition.

For another application of the third principle, consider scrub-jays who

[11] Some commentators suggested that this principle requires an ability to remember particular encounterings and their temporal order. There is some evidence that some nonhumans possess this ability (e.g. Clayton et al. 2003; Griffiths et al. 1999), but our proposal does not depend on this. To track another agent's most recent encounter, it is not necessary to remember all encounters and their temporal order; forgetting all but the most recent will do.

[12] The 'other direction' is required (in conjunction with the fourth principle, below) for explaining infants' success on false belief tasks where information about another's beliefs is provided not by what she sees but by what she does (as in Träuble et al. 2010).

strategically re-cache food depending on who saw what. In one experiment, scrub-jays cached some food in the presence of Competitor A and then cached more food in the presence of Competitor B. Later they had an opportunity to recover food in the presence of Competitor A. The scrub-jays preferred to recover and re-cache food cached in the presence of Competitor A, leaving untouched food cached in the presence of the absent competitor (Clayton et al. 2007, pp. 517-9). This strategy reduces the chances of Competitor A knowing where any of the scrub-jays' food is. What might explain this sensitivity to who saw what in choosing which food to recover? Clayton and colleagues postulate a capacity to attribute knowledge or 'informational states' (Clayton et al. 2007). They are opposed by Penn and colleagues who postulate that scrub-jays are not ascribing knowledge but using rules such as 'Try to re-cache food in a site different from the one where it was cached when the competitor was present' (Penn et al. 2008; Penn & Povinelli 2007). As an alternative to both proposals, the scrub-jays' sensitivity could be explained in terms of registration. Suppose scrub-jays understand that correct registration is a condition on successful goal-directed action. Then they may be trying to prevent competitors from stealing their cached food by means of preventing them from correctly registering its location.

A third application of the third principle is to infant pointing. Liszkowski et al. (2006) show that 12- and 18-month-olds point in order to provide relevant information to adults about the locations of objects. In their experiment, infants watch as an adult uses an object in some task. Then the adult appears to accidentally misplace it. Later, when the adult visibly needs that object to complete the task again, infants reliably point to it. Apparently infants point not in order to get the object or to share interest in it, but to enable the adult to complete a task. The authors suggest that this could be explained either by supposing that infants understand what the adult does not know, or that they understand what the adult is not attending to (Liszkowski et al. 2006, p. 185). As knowledge and attention are both complex psychological notions, these suggestions raise the hard question of what children might understand of them. Another possible explanation involves registration. Maybe the infants understand that correct registration is necessary for successful goal-directed action and that pointing is a way of generating correct registration. (None of these explanations bear on what is perhaps most interesting about these findings, that infants care to inform others and do so spontaneously. Minimal theory of mind as described here is at most a small part of a larger story.)

Note that have not argued for the truth of any hypothesis about what chimpanzees, scrub-jays or infants represent. As explained above, what motivates our enquiry is not that there is already evidence for hypotheses about particular subjects' minimal theory of mind cognition: it is that, given the limited evidence for competing hypotheses and the objections they face, new

hypotheses are worth formulating and testing. And in later sections we will explain how hypotheses about minimal theory of mind generate testable predictions, enabling them to be empirically distinguished from competing hypotheses. For now our focus is the question of what a subject could represent that would enable her to track track perceptions, knowledge states and beliefs in ways measured by various experimental paradigms.

4.4. Fourth principle

So far Lucky thinks of correct registration as a condition for the success of goal-directed action. This does not tell her anything about what happens if the condition is not met. In particular it tells her nothing about how an agent will act when she registers an object incorrectly. The fourth principle involves a switch from thinking of registration as a success condition to thinking of it as a causal factor. This principle states that when an agent performs a goal-directed action with a goal that specifies a particular object, the agent will act as if the object were in the location she registers it in.

Now that Lucky understands registration as a factor influencing action it can serve her as a proxy for false belief. Just as, in a limited but useful range of cases, you can track food sources' toxicities by representing their odours and prospective sexual partners' virtues by representing their plumage, so also you can track beliefs by representing registrations.

Applications of the fourth principle therefore include Onishi and Baillargeon's (2005) false belief task. Subjects are shown an adult observer who is present while a piece of melon is placed in one box. In the critical condition, the adult observer is then absent while the melon moves to another box. Comparative looking times indicate that the subjects, who are 14-month-old infants, expect that the adult will reach into the box not containing the melon.[13] The authors explain this finding by hypothesizing that the infants are ascribing beliefs about the melon's location to the adults (Onishi & Baillargeon 2005, p. 257). Alternatively, the findings could be explained on the hypothesis that they are tracking registration as a cause of action.

4.5. Extensions and variations

With the fourth principle we completed the construction of a minimal theory of mind capable of underwriting success on some false belief tasks. We stop here because false belief tasks are often taken to be an acid test for theory of mind. But of course additonal principles can be added accommodate further theory of mind abilities. For instance, further principles might extend

[13] This finding is supported by a growing body of related research (including Luo & Baillargeon 2007; Scott & Baillargeon 2010; Song et al. 2008; Southgate et al. 2007; Surian et al. 2007).

the definition of registration. Registration was been defined as a relation between agents, objects and locations. This definition could be extended to include other types of property as relata in addition to locations. Further notions, such as a relational proxy for tracking desires, could also be added. These and other modifications would enable hypotheses about minimal theory of mind cognition to explain a wider range of theory of mind abilities.[14]

Variations on the principles are also possible. To illustrate, scrub-jays' caching strategies do not obviously call for an ability to track relations to particular food items: their ultimate aim probably isn't to prevent others pilfering particular worms but to prevent them from pilfering any worms at all. Accordingly scrub-jays' protective caching strategies may involve tracking relations between agents, locations and food types rather than particular objects. In general it is possible that there is variation across species in what individuals represent that enables them to track perceptions, knowledge states and beliefs. If we attempted to characterise theory of mind cognition using only adult human commonsense psychological notions of perception, knowledge and belief, it would be hard to make systematic sense of the possibility of such variation. These commonsense psychological notions are not easy to take apart because they resemble the decision-theoretic notion of expected utility in this respect: they are characterised in part by their roles in a system of explanation (Davidson 1985, 1999b). One virtue of the minimal theory of mind construction is that it enables us to make systematic sense of the possibility of variations.

4.6. But is it theory of mind cognition?

The term 'theory of mind' has been used in several apparently distinct ways (Whiten 1994). On some definitions, all theory of mind cognition involves metarepresentation.[15] Minimal theory of mind involves representing goals to which actions are directed, encounterings and registrations, none of which are are representations. So if we accepted this definition, what we have constructed would not count as a form of theory of mind cognition just because no metarepresentation is involved.

On another influential definition, theory of mind cognition begins when

[14] Infants and two-year-olds' theory of mind abilities involve more than tracking false beliefs about location. To illustrate, Song et al. (2008) show that infants can use communicated information in predicting actions based on false belief, He et al. (2011) show that 2.5-year-olds can succeed on false belief tasks involving unexpected contents rather than location properties, and Scott et al. (2010) show that 18-month olds are sensitive to inferences that others are likely to make.

[15] See Perner (1991). This view can seem obviously correct if it is assumed that mental states are all representations. However, there reasons to doubt this assumption (see, for instance, Campbell 2002).

subjects ascribe states which function as variables intervening between environmental or behavioural inputs and behavioural outputs, and which play some roles characteristic of mental states (Whiten 1996; Penn & Povinelli 2007, p. 732). On this definition, the endpoint in our construction (but no earlier point) does count as theory of mind cognition because registrations are intermediate variables and play a subset of the causal roles characteristic of belief. For registrations, like beliefs, are characterised by principles generalising across all goal-directed actions, can be assigned correctness conditions and causally influence actions. (Of course they also differ from beliefs in many ways: they are not propositional attitudes, they cannot involve reference to never-existent objects and they are not subject to the norms that arguably characterise belief.) We use the label 'minimal theory of mind' because, relative to this definition, the construction describes a minimally elaborate form of theory of mind cognition, one that does not the entail cognitive and conceptual demands associated with representing perceptions, knowledge states and beliefs as such but is capable of grounding theory of mind abilities that generalise across goal-directed actions and are sufficient for systematic success on some false belief tasks.

In what follows we show that minimal theory of mind cognition has clear signature limits. These limits generate predictions capable of distinguishing hypotheses about minimal theory of mind both from hypotheses about representations of nonintentional behaviour only and also from hypotheses about full-blown theory of mind cognition.

5. Limits: how to distinguish minimal from full-blown theory of mind cognition

How could we distinguish minimal from full-blown theory of mind cognition experimentally? The point of minimal theory of mind is to enable agents to fake it—that is, to act as if they were reasoning about propositional attitudes, within limits. Where a task goes beyond these limits, we can be sure an agent is not using minimal theory of mind only.

Some limits on minimal theory of mind cognition arise from the fact that the theory makes use of objects and their relations to agents, rather than representations of objects, to predict others' behaviours. This means that false beliefs involving quantification or identity cannot be tracked by representing registrations. To see why not, consider the following inference:

(1) Mitch believes that Charly is in Baltimore.

(2) Charly is Samantha.

Therefore:

(3) Mitch believes that Samantha is in Baltimore.

On almost any account of belief, this inference is not valid (Frege 1948, pp. 214-5). Its central role in a popular film (Harlin 1996) indicates that human adults typically appreciate that this inference is not valid. Contrast the above inference with the corresponding inference in the case of registration:

(1′) Mitch registers <Charly, Baltimore>

(2) Charly is Samantha.

Therefore:

(3′) Mitch registers <Samantha, Baltimore>

This inference from (1′) and (2) to (3′) is logically valid. It is valid because registration is a relation to objects. We can compare registration with other relations like being left of something. If Charly is Samantha (whether you know it or not), then anyone who is left of Charly is left of Samantha; similarly for registering Charly's location.

This formal difference between belief and registration entails a limit on minimal theory of mind cognition. Consider Lucky who tracks beliefs by means of representing registrations only and is unable to represent beliefs. Lucky should have no problem predicting actions based on false beliefs about the locations of objects but she should encounter difficulties in predicting actions based on beliefs essentially involving mistakes about identity. In particular, Lucky should not be able to understand why, when Mitch registers <Charly, Baltimore>, he continues searching for Samantha.[16] For to register <Charly, Baltimore> is the same thing as registering <Samantha, Baltimore>. And Lucky should be equally at a loss when those she observes mistakenly believe that two distinct people are identical. By contrast, subjects who can represent beliefs as such should have no special problem with false beliefs essentially involving identity. This is how mistakes about the identities of objects can be used to distinguish minimal from full-blown theory of mind cognition.[17]

How could this be exploited experimentally? One paradigm suitable for infants would involve a puppet which has a concave face and body so that, unusually, it looks very different when seen from different angles. Viewed from one angle, the puppet looks like one person (Charly); viewed from another angle, the puppet looks like another person (Samantha); and viewed head-on both aspects of its appearance become visible simultaneously. Crucially, it is possible to see either aspect without having reason to suppose

[16] This assumes that Lucky herself knows that Charly is Samantha. To ease exposition we assume throughout that Lucky has no false beliefs involving identity.

[17] Related points about quantification entail further testable distinctions. For instance, minimal theory of mind should make it impossible to track beliefs whose contents essentially involve *most* objects having a certain property. It is beyond the scope of this paper to elaborate on these predictions.

that the puppet would look so different from another angle. Subjects sit at a table opposite the protagonist. There is a screen in the middle of the table which blocks each person's view of the other side of the table. In the critical condition, the puppet is initially on the subjects' side of the screen (see Figure 2 on the following page). In the first scene, the puppet emerges from the screen so that both observers can see it. Only one aspect of the puppet is revealed to each; the protagonist sees the Charly-aspect, the subject sees the other aspect. In the second scene the puppet returns to the subjects' side of the screen; she then rotates. The subject but not the protagonist is thereby shown that the face and body have two aspects. Finally, in scene three, the puppet emerges from the screen. The puppet again reveals only one aspect of its face and body to each observer but this time the protagonist sees the Samantha-aspect and the subject sees the other aspect. At this point the puppet leaves the stage altogether. Subjects can see that there is nothing on their side of screen, but the protagonist cannot see this. Critically, this means that there is reason for the protagonist to expect there still to be a puppet (Charly) behind the screen. So when, finally, the protagonist reaches around the screen, this action makes sense: though there is actually nothing behind the screen, the protagonist has reason to expect there is a puppet. But for subjects to appreciate that the protagonist has this expectation, and to make sense of her reaching action, they would have to ascribe a false belief about identity.[18] Subjects' lack of surprise at the protagonist's reaching around the screen (as compared to a control condition which differs from this one only in that the protagonist manifestly knows that the two aspects of the puppet's face and body are aspects of a single puppet) would therefore be evidence that they can track false beliefs about identity. This would be evidence for full-blown, not minimal, theory of mind.

So far we have discussed mistaken beliefs about identity where the truth of the belief would require there to be more individuals in the world (Charly and Samantha; a teacher and an assassin) than there actually are. The converse is also possible (see, e.g., Coen & Coen 1998). Here the mistake is to have beliefs whose truth would require there to be fewer individuals in the world than there actually are. The main condition of a possible experiment is outlined in Figure 3 on page 23. In this experiment there are two perceptually indistinguishable balls. Whereas subjects can see both balls simultaneously, a screen ensures that the protagonist never sees both at once. The movements of the two balls are timed in such a way that what the protagonist sees indicates the existence of a single ball.[19] When the protagonist sees a

[18] To ensure this, the design must be such that subjects cannot readily categorise the puppet's two appearances and such that subjects do not use verbal labels for the two aspects. Otherwise it would be possible for subjects to infer the protagonist's expectation by reasoning about types of object.

[19] Song and Baillargeon (2008) provide evidence that 15-month-old infants are able to pre-

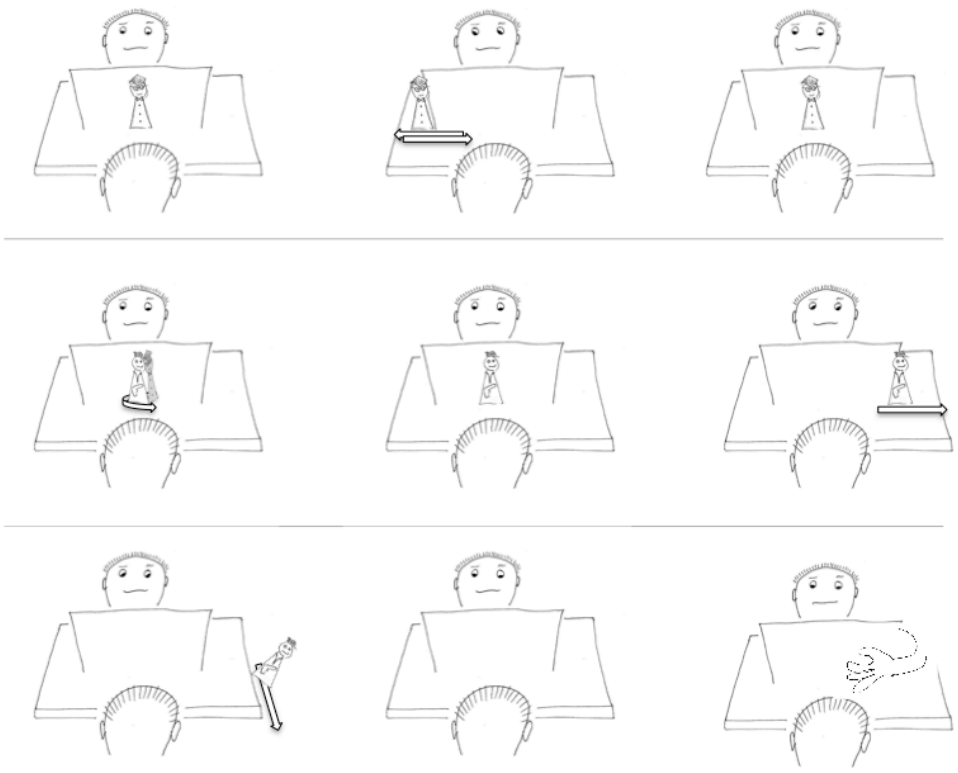

Figure 2: How to test whether infants can ascribe false beliefs about identity. The subject is in the foreground with only the back of her head visible; the protagonist sits opposite, facing the subject. A screen conceals some of the action from the protagonist.

ball appear from behind the screen and leave the scene (panels 5-7 in Figure 5), this provides a reason for the protagonist to believe that there is nothing behind the screen. Infants could only be sensitive to this belief if they were able to track beliefs involving identity. Sensitivity to this belief might be manifested by comparing this condition with a modified condition in which the protagonist manifestly knows that there are two balls. If subjects show more surprise when the protagonist reaches around the screen (say) in the original condition than in the modified condition, this too would be evidence for full-blown over minimal theory of mind cognition.

Given that this analysis, it is notable that one study already claims to have found positive evidence that infants can ascribe beliefs about identity. Scott and Baillargeon (2010) have already conducted experiments which aim to show that 18-month-old infants are able to represent beliefs involving mistakes about 'which particular object *token*' another is facing rather than only about 'which *type* of object' is present (p. 1179). However, while their aim is clearly relevant, we shall show that these experiments do not succeed in testing whether infants ascribe beliefs about identity. In essence, their paradigm involves two types of object, a divisible penguin which affords hiding a key and an indivisible penguin which lacks this affordance. The divisible penguin has two states, divided and whole. When whole it is visually indistinguishable from the indivisible penguin which does not afford hiding a key. Infants sit opposite a protagonist. Familiarisation trials provide information that the protagonist can expect there to be two penguins in the scene at all times, one of each type; these trials also provide information that the protagonist can expect the divisible penguin to appear in its divided state. In the critical condition, the scene contains an uncovered and a covered penguin. The uncovered penguin is divisible but, because it is in the whole state, it might easily be taken to be indivisible. The covered penguin is indivisible. The findings suggest that infants know that the uncovered penguin is divisible but expect the protagonist to believe that the uncovered penguin is indivisible and to infer that the covered penguin is divisible. Does this pattern of knowledge and expectation require that the infants have attributed a false belief about identity? This is not necessary. Suppose that infants attributed the following reasoning to the protagonist: this penguin visible on my right is *an* indivisible penguin; *a* divisible penguin is always present; therefore there is *a* divisible penguin under the cover. This reasoning concerns only the types of objects present and does not involve identity. Yet it is sufficient (together with some further inferences about the protagonists'

dict actions on the basis of how things appear to observers who are ignorant of their true nature. Schmelz et al. (2011) show that chimpanzees also make inferences about the appearances of things in predicting behaviours. These and other findings (e.g. Luo & Beck 2010) suggest that infants and perhaps others are generally sensitive to facts about the ways things look in tracking others' beliefs.

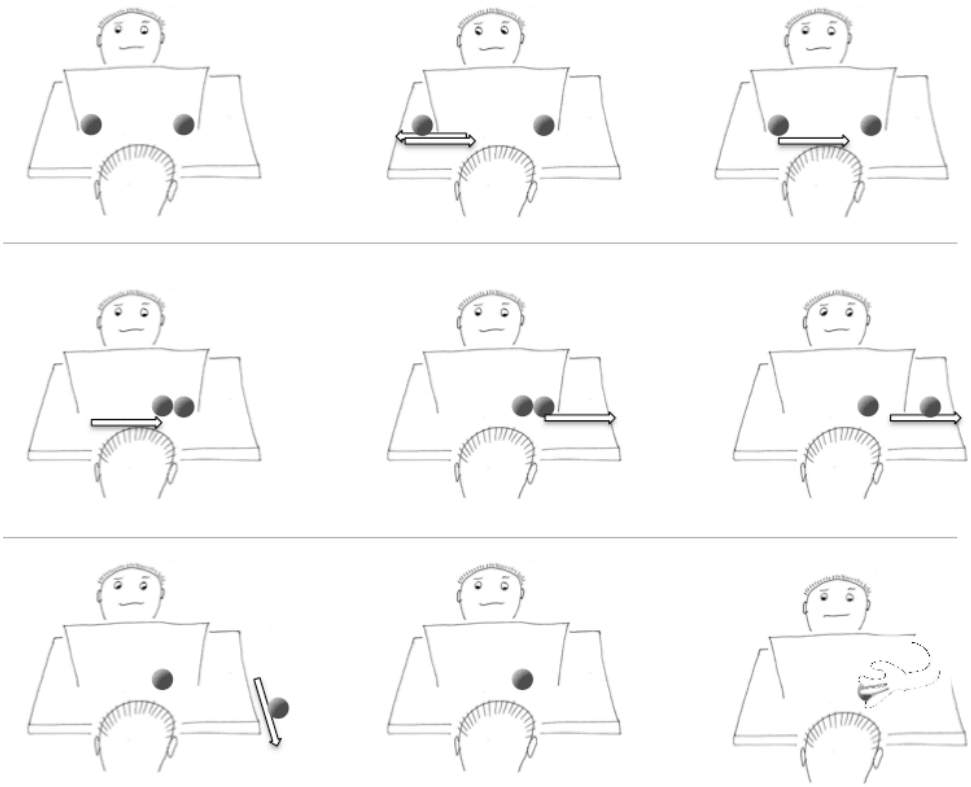

Figure 3: How to test whether infants can ascribe false beliefs about identity where the mistake involves belief whose truth would entail that there were fewer objects than there in fact are. As in Figure 2, the subject is in the foreground with only the back of her head visible; the protagonist sits opposite, facing the subject. A screen conceals some of the action from the protagonist.

goal) for success. So while Scott and Baillargeon's (2010) findings are interesting in many respects, their implementation does not provide evidence that infants ascribe beliefs about identity.

In this section we have explained ways of distinguishing minimal theory of mind cognition from its full-blown counterpart. Where subjects are able to track beliefs about identity we know they are not relying only on minimal theory of mind; and, conversely, where subjects are able to track beliefs about objects' properties but not about identity, we can infer that they are not using full-blown theory of mind. We next consider how minimal theory of mind might be distinguished experimentally from strategies which involve representing nonintentional behaviour only.

6. Limits: how to distinguishing minimal theory of mind from behavioural strategies

What could show that subjects are not solving a task by representing nonintentional behaviour only but are using minimal theory of mind or something stronger? We shall use the term 'behavioural strategy' to refer to processes underpinning theory of mind abilities which involve representing nonintentional behaviours only.

An obstacle to distinguishing minimal theory of mind from behavioural strategies is that there is no in-principle limit to the complexity of the predictions which behavioural strategies might support. The behaviour-representing counterpart of Laplace's demon knows our entire behavioural histories, has perfect knowledge of the regularities governing object-directed actions and is not limited by scarce cognitive resources. This demon can readily predict our future behaviours without any theory of mind.

One way to avoid the problem of the behaviour-representing demon is to set tasks involving explanation rather than prediction (compare Salmon 1998, p. Chapter 8). Subjects observe a standard false belief scenario in which a character, Maxi, first searches for an object in the wrong location. The test question is why Maxi searched as he did (see e.g. Moses & Flavell 1990; Wimmer & Mayringer 1998). Where subjects' answers include reference to incorrect registration or a related notion (which might, of course, be verbalised in terms of what an agent 'thinks'), they have at least minimal theory of mind cognition. Of course, this only works where subjects can be asked 'why' questions.

For other subjects we need an indirect measure. If subjects sometimes represent encounters or perceptions as such, then representations of others' encounters may interfere with expectations or judgements about reality or with judgements about one's own perceptions and beliefs (so-called 'altercentric interference', Samson et al. 2010). For example, representing a

protagonists' encounter with an object at a particular location may interfere with expectations about the object's actual location. Accordingly, evidence that there is such interference when in a situation where others have different perceptions or beliefs would provide indirect evidence for theory of mind cognition (minimal or full-blown) over behavioural strategies, for such interference would be difficult to explain if subjects were representing behaviours only. Evidence showing this kind of altercentric interference has recently been found in human adults (Qureshi et al. 2010b; Samson et al. 2010). These studies show that adults' judgements about the number of objects they could see in a visual scene were slower and more error-prone when the scene contained an irrelevant other whose perspective was different, suggesting that another's perspective can cause altercentric interfere.

Of course it was never in question that adults represent encounters or perceptions and not only behaviours; the relevance of these studies is that they demonstrate that altercentric interference can be used as an indirect measure for distinguishing behavioural strategies from minimal theory of mind cognition. Kovács et al. (2010) report evidence of possible altercentric interference in infants' expectations about the presence of an object. Their study suggests that infants' surprise when their expectations about an object's presence or absence are violated is modulated by facts about another's expectations about that object's location. Infants looked for less time at the unexpected presence of an object behind a screen when another observer incorrectly expected it to be there than when the observed did not expect it to be there. These findings may be evidence of altercentric interference,[20] which would be consistent with either minimal or full-blown theory of mind but difficult to explain in a principled way if subjects are using behavioural strategies only.

A related way to distinguish behavioural strategies from minimal theory of mind or something stronger is by means of self-other inference. Emery and Clayton (2001) divided scrub-jays into those who had previously pilfered food and those who had never pilfered. All subjects then cached some food in private and cached some more food while being observed. When later given privacy, the pilferers re-cached the food they had earlier cached in the presence of the observer but did not re-cache the food they had cached in private. The non-pilferers did not show this pattern of behaviour and did little re-caching.[21] This is difficult to reconcile with the hypothesis that scrub-

[20] An anonymous referee pointed out that this is not the only possible interpretation. We do not mean to suggest that a single study provides decisive evidence for altercentric interference, only that in the wider context of research on infant theory of mind it provides an indication and constitutes progress on the question of how altercentric interference could be measured.

[21] Similar results might be found if this paradigm were adapted for human infants (see Sommerville & Woodward 2005).

jays rely entirely on behavioural strategies. By contrast, the hypothesis that scrub-jays have at least minimal theory of mind and so are able to represent goal-directed actions as such allows us to suppose that engaging in pilfering alerts them to the possibility that others may do the same.[22]

Another route to distinguishing minimal theory of mind cognition from behavioural strategies derives from an analogy with speech perception. It is well established that phoneme recognition is facilitated by lexical information (Samuel 2001). Similarly, for subjects with (at least) minimal theory of mind cognition, there is evidence that identification of significant segments of behaviour is facilitated by information about goals. For instance, Zacks (2004) varied the availability of information about goals while holding motion features of behaviour constant. He found that subjects given information about goals segmented behaviour differently from subjects not given such information. Of course for Zacks it was not in question that his subjects could represent goals; the issue was how their representation of goals interacted with their representations of behaviour. But with modification the paradigm could be used to test participants who are not already known to be capable of representing goals. If making information about goals, encounterings or registrations available alters how subjects segment behaviours, this would be evidence that they are not relying on behavioural strategies only.

In short, a variety of routes to distinguishing behavioural strategies from minimal theory of mind are available. Some of these involve established paradigms, others adapt ideas from related research. A decisive argument either way would require converging evidence from multiple paradigms (Premack & Woodruff 1978, p. 622).

7. Conclusion

Abilities to track[23] perceptions and beliefs are sometimes but not always automatic in human adults (or so we argued in Section 2). But, as a variety of theoretical and empirical considerations indicate, representing beliefs and other propositional attitudes as such is associated with demands on working memory, inhibition or attention that are incompatible with automaticity. This motivated asking what someone could represent that would enable her to track perceptions, knowledge states and beliefs without meeting these cognitive demands.

---

[22] In later discussions these authors claim that the pilferers may be projecting their own experiences onto others in the sense that they are using 'their own experiences to infer experiences in others' (Emery & Clayton 2007, p. 81). This claim goes beyond our suggestion, which involves goals rather than experiences and salience rather than projection.

[23] The notion of *tracking* was defined on page 3; see also the definitions of *theory of mind ability* (on page 2) and *theory of mind cognition* (on page 3).

Further motivation for asking this question comes from evidence that theory of mind abilities are widespread, occurring not only in human adults but also in infants, chimpanzees and scrub-jays at least (see Section 3). By contrast full-blown theory of mind cognition may be comparatively rare thanks to the conceptual and cognitive demands associated with representing beliefs and other propositional attitudes as such. If so it is useful to identify what else someone could represent that would enable her to track, in a limited but useful range of situations, perceptions, knowledge states and beliefs, including false beliefs.

To answer this question we constructed a minimal theory of mind. The construction is rich enough to explain systematic success on tasks held to be acid tests for theory of mind cognition including many false belief tasks. Where minimal theory of mind must break down is in cases involving quantification or mistakes about identity (see Section 5). Because such cases require full-blown theory of mind, it is possible to distinguish whether an individual's performance on a particular task involves minimal or full-blown theory of mind cognition.

The central notions for minimal theory of mind are encountering and registration. We characterised these by their structures, acquisition conditions and roles in producing goal-directed actions (see Section 4). Encountering and registration serve as non-representational proxies for perception and belief in this sense: in a limited but useful range of everyday circumstances, agents perceive an object just when they encounter it and they believe that an object has a given property just when they register it having that property. But encountering differs from perception in not constitutively involving appearances, modalities or the possibility of illusion and in not being constitutively linked to reasons, knowledge or informational states. Similarly, registration differs from belief in lacking representational content and the possibility of embedding, in not interacting in arbitrarily complex ways with other states, and in the simplicity of its parameter-setting role in causing actions. These differences ensure that representing encountering or registration need not involve meeting conceptual and cognitive demands associated with representing perceptions, beliefs and other propositional attitudes as such.

The novelty of our constructive approach lies in several features. It does not rely directly on everyday psychological concepts, whose exact nature is a source of controversy. Nor does it rely on infants or non-human animals holding theoretical commitment to simplified versions of these concepts (contrast Gopnik & Meltzoff 1997; Perner 1991; Wellman & Bartsch 1994 and Wellman et al. 2000). Importantly, we do not assume that minimal theory of mind develops into full-blown theory of mind in humans. It may instead remain distinct, supporting cognitively efficient theory of mind across the lifespan (see Samson et al. 2010; Schneider et al. 2011). The construction

of minimal theory of mind is systematic enough to generate testable predictions distinguishing it from both behavioural stratetgies and full-blown theory of mind cognition (see Sections 5 and 6). Our construction makes detailed sense of the notion that there are degrees of theory of mind cognition. (This is a virtue because while is widely recognised that degrees of theory of mind cognition are needed (e.g. Bartsch & Wellman 1995; Whiten 1994), there have been few detailed attempts to make systematic sense of this possibility.) It also pushes much further than earlier work the boundaries of what can be achieved without full-blown theory of mind cognition; in particular, it explains how systematic success on a range of false belief tasks (but not those which essentially involve identity or quantification) is possible without representing beliefs or other propositional attitudes as such. Minimal theory of mind may be what enables those with limited cognitive resources or little conceptual sophistication, such as infants, chimpanzees, scrub-jays and human adults under load, to track others' perceptions, knowledge states and beliefs.

References

Apperly, I., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093–1108.

Apperly, I., Carroll, D., Samson, D., Humphreys, G., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217.

Apperly, I., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844.

Apperly, I., Samson, D., Chiavarino, C., & Humphreys, G. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, *16*(10), 1773–1784.

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

Apperly, I. A., Samson, D., & W., H. G. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, *45*(1), 190–201.

Astington, J. & Baird, J. A. (Eds.). (2005). *Why Language Matters for Theory of Mind*. Oxford: Oxford University Press.

Back, E. & Apperly, I. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407.

Baldwin, D. & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, *5*(4), 171–178.

Bartsch, K. & Wellman, H. M. (1995). *Children talk about the mind*. New York ; Oxford: Oxford University Press.

Birch, S. A. J. & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386.

Bugnyar, T., Stöwe, M., & Heinrich, B. (2007). The ontogeny of caching in ravens, corvus corax. *Animal Behaviour*, *74*(4), 757–767.

Bull, R., Phillips, L., & Conway, C. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition*, *107*(2), 663–672.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342.

Byrne, R. W. (1999). Imitation without intentionality. using string parsing to copy the organization of behaviour. *Animal Cognition*, *2*(2), 63–72.

Call, J. & Tomasello, M. (2005). What chimpanzees know about seeing revisited: An explanation of the third kind. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention:Communication and other Minds* (pp. 45–64). Oxford: Oxford University Press.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Carlson-Radvansky, L. A. & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, *33*, 646–671.

Casler, K. & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition*, *103*, 120–130.

Cassam, Q. (2008). Knowledge, perception and analysis. *South African Journal of Philosophy*, *27*(3), 48, 36–48, 36.

Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat Rev Neurosci*, *4*(8), 685–691.

Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. the western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B*, *362*, 507–552.

Clayton, N. S. & Emery, N. J. (2007). The social life of corvids. *Current Biology*, *17*(16), R652–R656–R652–R656.

Clements, W., Rustin, C., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. *Developmental Science*, *3*(1), 81–92.

Coen, J. & Coen, E. (1998). The big lebowski.

Csibra, G. & Gergely, G. (2007). Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78.

Dally, J. M., Emery, N. J., & Clayton, N. S. (2004). Cache protection strategies by western scrub-jays (aphelocoma californica): hiding food in the shade. *Proceedings of the Royal Society B: Biological Sciences*, *271*(0), S387–S390–S387–S390.

Davidson, D. (1980). Towards a unified theory of meaning and action. *Grazer Philosophische Studien*, *11*, 1–12.

Davidson, D. (1985). A new basis for decision theory. *Theory and Decision*, *18*, 87–98.

Davidson, D. (1990). The structure and content of truth. *The Journal of Philosophy*, *87*(6), 279–328.

Davidson, D. (1999a). The emergence of thought. *Erkenntnis*, *51*, 7–17.

Davidson, D. (1999b). The emergence of thought. *Erkenntnis*, *51*, 7–17.

Doherty, M. & Perner, J. (1998). Metalinguistic awareness and theory of mind: just two words for the same thing? *Cognitive Development*, *13*, 279–305.

Doherty, M. J. (2006). The development of mentalistic gaze understanding. *Infant and Child Development*, *15*, 179–186.

Dretske, F. (1969). *Seeing and Knowing*. London: Routledge.

Dretske, F. (1988). *Explaining Behavior*. Cambridge, Mass.: MIT Press.

Emery, N. J. & Clayton, N. S. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature*, *414*, 443–446.

Emery, N. J. & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, *306*(5703), 1903–1907.

Emery, N. J. & Clayton, N. S. (2007). How to build a scrub-jay that reads minds. In S. Itakura & K. Fujita (Eds.), *Origins of the Social Mind: Evolutionary and Developmental Perspectives*. Tokyo: Springer.

Ferguson, H. J. & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*(1), 257–263.

Frege, G. (1948). Sense and reference. *The Philosophical Review*, *57*(3), 209–230.

Frischen, A., Loach, D., & Tipper, S. (2009). Seeing the world through another person's eyes: Simulating selective attention via action observation. *Cognition*, *111*, 212–8.

German, T. P. & Defeyter, M. A. (2000). Immunity to functional fixedness in young children. *Psychonomic Bulletin and Review*, *7*(4), 707–12.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Gomez, J.-C. (2007). Pointing behaviors in apes and human infants: A balanced interpretation. *Child Development*, *78*(3), 729–734.

Gopnik, A. & Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Learning, development, and conceptual change. Cambridge, Mass.: MIT Press.

Griffiths, D., Dickinson, A., & Clayton, N. (1999). Episodic memory: what can animals remember about their past? *Trends in Cognitive Sciences*, *3*(2), 74–80.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151.

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition*, *101*(3), 495–514.

Harlin, R. (1996). The long kiss goodnight.

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Developmental Science*, *14*(2), 292–305.

Hughes, C., Fujisawa, K. K., Ensor, R., Lecce, S., & Marfleet, R. (2006). Cooperation and conversations about the mind: A study of individual differences in 2-year-olds and their siblings. *British Journal of Developmental Psychology*, *24*(1), 53–72.

Hughes, C. & Leekam, S. (2004). What are the links between theory of mind and social relations? review, reflections and new directions for studies of typical and atypical development. *Social Development*, *13*(4), 590–619.

Kelemen, D. (1999). Function, goals and intention: children's teleological reasoning about objects. *Trends in Cognitive Sciences*, *3*(12), 461–468.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Knudsen, B. & Liszkowski, U. (2012). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, *forthcoming*.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830 –1834.

LeFevre, J., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory & Cognition*, *16*(1), 45–53.

Leslie, A. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*(10), 459–62.

Leslie, A., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, *50*, 45–85.

Lewis, C. (1994). Episodes, events, and narratives in the child's understanding of mind. In C. Lewis & P. Mitchell (Eds.), *Children's Early Understanding of Mind: origins and development*. Hove: Erlbaum.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). Twelve- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*(2), 173–187.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739.

Luo, Y. & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512.

Luo, Y. & Beck, W. (2010). Do you see what i see? infants' reasoning about others' incomplete perceptions. *Developmental Science*, *13*(1), 134–142.

Marr, D. (1982). *Vision : a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Martin, M. G. F. (2002). The transparency of experience. *Mind & Language*, *17*(4), 376–425.

McKinnon, M. C. & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.

Michotte, A. (1946 [1963]). *The Perception of Causality*. London: Meuthen.

Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.

Moses, L. & Flavell, J. (1990). Inferring false beliefs from actions and reactions. *Child Development*, *61*, 929–945.

Neander, K. (1991). The teleological notion of a function. *Australasian Journal of Philosophy and Phenomenological Research*, *69*, 454–68.

Newtson, D. & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–50.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659–677.

Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(8), 255–258.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), 109–130.

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 B*, *362*(1480), 731–744.

Perner, J. (1990). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing Theories of Mind* (pp. 141–172). Cambridge University Press.

Perner, J. (1991). *The Representational Mind*. Brighton: Harvester.

Perner, J. & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*(9), 337–344.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development*, *73*(3), 752–767.

Perner, J. & Ruffman, T. (2005). Infant's insight into the mind: How deep? *Science*, *308*, 214–6.

Povinelli, D. J. & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, *7*(4), 157–160.

Povinelli, D. J. & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, *19*(1), 1–28.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526.

Price, C. (2001). *Functions in Mind.* Oxford: Clarendon Press.

Qureshi, A., Apperly, I., & Samson, D. (2010a). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.

Qureshi, A., Apperly, I. A., & Samson, D. (2010b). Executive function is necessary for perspective-selection, not level 1 visual-perspective calculation: evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.

Rakoczy, H., Tomasello, M., & Striano, T. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. In L. L. Namy (Ed.), *Symbol use and symbolic representation: Developmental and comparative perspectives* (pp. 69–97). Mahwah, NJ: Lawrence Erlbaum.

Ruffman, T. & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences*, *9*(10), 462–3.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, *16*, 233–253.

Sabbagh, M. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, *77*(4), 1034–1049.

Salmon, W. (1998). *Causality and Explanation.* Oxford: Oxford University Press.

Samson, D., Apperly, I., Kathirgamanathan, U., & Humphreys, G. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, *128*(5), 1102–1111.

Samson, D., Apperly, I. A., Braithwaite, J. J., & Andrews, B. (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.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science: A Journal of the American Psychological Society / APS*, *12*(4), 348–351.

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–113.

Schmelz, M., Call, J., & Tomasello, M. (forthcoming 2011). Chimpanzees know that others make inferences. *Proceedings of the National Academy of Sciences*.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2011). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *advance online*.

Scholl, B. J. & Nakayama, K. (2004). Illusory causal crescents: Misperceived spatial relations due to perceived causality. *Perception*, *33*, 455–469.

Scott, R. M. & Baillargeon, R. (2010). Which penguin is this? attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172–1196.

Scott, R. M., Baillargeon, R., Song, H.-j., & Leslie, A. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*(4), 366–395.

Slaughter, V. & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, *67*, 2967–2988.

Sommerville, J. A. & Woodward, A. L. (2005). Pulling out the intentional structure of action: the relation between action processing and action production in infancy. *Cognition*, *95*(1), 1–30.

Song, H.-j. & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology. Vol. 44(6)*, *44*(6), 1789–1795.

Song, H.-j., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? psychological reasoning in 18-month-old infants. *Cognition*, *109*(3), 295–315.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, *18*(7), 587–592.

Suddendorf, T. & Whiten, A. (2003). Reinterpreting the mentality of apes. In J. Fitness & K. Sterelny (Eds.), *From Mating to Mentality: Evaluating Evolutionary Psychology* (pp. 173–196). Psychology Press.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586.

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2011). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*.

Taylor, C. (1964). *The Explanation of Behaviour*. London: Routledge.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.

Tomasello, M. & Call, J. (2005). Do chimpanzees know what others see–or only what they are looking at? In S. Hurley & M. Nudds (Eds.), *Rational Animals* (pp. 371–84). Oxford: Oxford Univesity Press.

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, *15*(4), 434–444.

Trick, L. & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? a limited-capacity preattentive stage in vision. *Psychological review*, *101*(1), 80.

Tye, M. (2002). Representationalism and the transparency of experience. *Nôus*, *36*(1), 137–151.

Vonk, J. & Povinelli, D. J. (2006). Similarity and difference in the conceptual systems of primates: The unobservability hypothesis. In E. Wasserman & T. Zentall (Eds.), *Comparative Cognition: Experimental Explorations of Animal Intelligence* (pp. 363–387). Oxford: Oxford University Press.

Wellman, H. & Bartsch, K. (1994). Before belief: children's early psychological theory.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, *72*(3), 655–684.

Wellman, H., Phillips, A. T., & Rodriguez, T. (2000). Young children's understanding of perception, desire, and emotion. *Child Development*, *71*(4), 895–912.

Whiten, A. (1994). Grades of mindreading. In C. Lewis & P. Mitchell (Eds.), *Children's Early Understanding of Mind* (pp. 47–70). Hove: Erlbaum.

Whiten, A. (1996). When does smart behaviour-reading become mind-reading? In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind* (pp. 277–292). Cambridge: Cambridge University Press.

Wimmer, H. & Mayringer, H. (1998). False belief understanding in young children: Explanations do not develop before predictions. *International Journal of Behavioral Development*, *22*(2), 403–422.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Wright, L. (1976). *Teleological Explanations*. Berkeley: University of California Press.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008.