Chapter 3

Teleology: Belief as perspective

Johannes Roessler and Josef Perner

A fundamental question in recent "theory of mind" research is how to interpret a seemingly robust dissociation between young children's performance on different kinds of tests for false belief understanding. 3-year-olds' poor performance on classical, "direct" false belief tasks is well-documented. Yet a range of "indirect" tests reveal sensitivity to agents' false beliefs in much younger children. It is natural to think that the two kinds of tests bring to light two kinds of understanding: "explicit" vs. "implicit" understanding. But how should we understand this distinction? And why should "implicit" understanding of false beliefs only be available in connection with "indirect" tests?

Our project in this chapter is to address these questions by further developing a hypothesis advanced elsewhere (Perner and Roessler, 2010). This is the hypothesis that young children are *teleologists*: they make sense of intentional actions in terms of justifying reasons provided by "worldly" facts (not by mental states). We begin by spelling out this account in more detail. We then argue that mastery of the concept of belief (or possession of an "explicit understanding" of belief) involves giving a twist to the teleological scheme of explanation. What is critical is the ability to engage in hypothetical or suppositional reasoning about justifying reasons. This account, we contend, is in competition with both a "theory theory" and a "simulation theory" of belief understanding (though it has some affinities with certain versions of the latter). In the final, fourth part of the chapter we bring the account to bear on the dissociation problem. The difference between "direct" and "indirect" tests, we argue, turns on whether successful performance requires understanding the normative underpinnings of the causal role of belief (as in direct tests) or merely requires a set of generalizations regarding the causes of behavior (as in indirect tests).

Teleological explanation

Why does the baker get up at 3 a.m.? Well, the bread needs to be ready by 6 to go to the supermarkets, and it takes that long to bake. This is a humble example of a teleological explanation: it makes the baker's unusual behavior intelligible not by appeal to his mental states, such as his desire to make bread etc, but in terms of the objective reason-giving facts of his situation. Our suggestion is that young children are teleologists. They predict, and perhaps explain, what someone will do on the basis of what it makes objective sense for her to do. This, we suggest, explains the following striking finding concerning young children's performance on false belief tests: far from answering the test question randomly, they systematically and adamantly give the wrong answer. The explanation is that they predict that the protagonist will do what he *ought to* do in order to attain his objective. For example, they will predict that in order to retrieve his chocolate Mistaken Max (Wimmer & Perner, 1983) will go to the cupboard (where he ought to go, as this is where the chocolate is to be found) rather than to the kitchen drawer (where he believes it is).

This needs some elaboration and qualification. You might say that there is a sense in which Max ought to go *to the drawer*. Given his false belief, surely it would be quite irrational for him to go the cupboard, where he has absolutely no reason to expect the chocolate. We agree. But the point is consistent with there *also* being a sense in which he ought to go to the cupboard: we have to recognize two kinds of practical "ought." Sometimes we are interested in whether someone ought to perform a certain action in the sense that there is *reason* for her to perform it. (An obvious context in which this question is to the fore is when you deliberate about what to do—i.e. reflect on what you have reason to do.) Sometimes we are interested in whether an agent *rationally* ought to perform a certain action, given her existing beliefs, aims and dispositions. Following Kolodny (2005) we will call these the "ought" of reasons vs. the "ought" of rationality. (Alternatively, one might put the contrast in terms of objective vs. subjective reasons.)

To make the distinction vivid, suppose we are advising Mistaken Max on what to do. From our vantage point as spectators of the story, the obvious recommendation is: "You ought to go to the cupboard—you have reason to: that's where the chocolate is." On the other hand, if Max remains firmly convinced that the chocolate is in the drawer, despite our best efforts to convince him otherwise, we might switch to a different kind of advice: "Given your belief, the rational thing for you to do is clearly to go to the drawer—that's what you ought to do.'[1] The first type of judgment is more prevalent in the context of advice and joint deliberation; the second type of judgment is more prevalent in the context of evaluating the rationality of an action. But they are both central and familiar elements of commonsense psychology.

Another way to bring out the distinction is to consider what happens when Mistaken Max (without the benefit of advice from us) goes to the kitchen table and opens the drawer. No doubt he'll be surprised. That's because he realizes that he was *wrong*: he thought there was a good reason for him to go to the drawer, but it now turns out that there wasn't. No chocolate—no reason. This is of course consistent with saying, as Max may find it comforting to say, that it was perfectly *rational* for him to go to the drawer.

The "ought" of rationality is often invoked in the context of action *explanation*. Those who emphasize the "rationalizing" nature of such explanations tend to have in mind that we explain intentional actions in terms of attitudes—centrally, beliefs and desires—that make it rational for the agent to perform the action. Our proposal is that young children think of intentional actions in a more simple-minded way: they predict and explain actions in terms of *reason*-giving facts, rather than *rationalizing* mental states.

But can such facts coherently be conceived as *causes*? As Davidson taught us, to explain why someone got up at 3 a.m., it is not enough to assemble considerations—"justifying reasons"—that show this to have been the right thing for him to do. What is required is a causal explanation (Davidson 1963). We grant the point. But we suggest that there is nothing incoherent in the idea that reason-giving "worldly" facts causally explain someone's actions. Note, first, that such facts yield reasons that can be "agent-specific." That the bread needs to be ready by 6, happily, has no implications as to when *you* ought to get up, but it gives the baker and his staff a reason to rise early. This is not because the reason in question is provided by the beliefs and desires of the relevant agents. Rather, agent-specificity is secured in this schema either by dint of the social roles of the agents (it is the baker's job to deliver the bread on time) or as a result of their practical abilities and opportunities for action (that Max needs his chocolate could give anyone a reason to help him get it if they are in a position to do so). You might say that without appeal to the agent's beliefs and

[1] These pieces of advice are modeled on Kolodny's examples of what he calls "objective" and "subjective" advice.

desires it's totally mysterious by what sorts of causal mechanisms the reason-giving facts impact on the agent's movements. But this does not impugn the *coherence* of the teleological schema, at least on what is sometimes called a "difference-making" approach to causal explanation (Woodward, 2011). To say that one fact causally explains another is to say that certain counterfactual conditionals hold: roughly, had there been some variation in respect of the first fact, there would have been a corresponding difference in the second fact. If the bread had not been needed until 7, the baker would have slept longer. You might still insist that, without some idea of the causal mechanisms involved, it would be quite irrational to make a causal judgment. Be that as it may (and the point is far from obvious), our claim is not that young children's simple-minded teleology is correct (or a model of rationality)—merely that it is a coherent explanatory (and indeed causal-explanatory) schema.

We have mentioned one piece of evidence in favor of the teleological analysis. If young children predict what people will do on the basis of teleological reasoning, it becomes comprehensible why they are so wedded to their predictions. It is not that they are unable to inhibit a prepotent response (in which case one would expect that once the mistake is pointed out, they realize what the correct answer is). Rather, their predictions are based on sound reasoning! People normally do what it makes sense for them to do. From a teleological point of view, what it makes sense for people to do depends on their objective circumstances—the relevant evaluative and instrumental facts. In other words, young children subscribe to a rather austere version of the "principle of charity," enjoining them to assume that people do what they have reason to do.[2]

Admittedly, the point hardly amounts to an open-and-shut case for teleology. Our aim here, though, is not to undertake a comprehensive review of the evidence. (See Perner and Roessler, 2010, for more detailed discussion.) Rather we want to argue that the teleological account provides an illuminating perspective on two vexed (and we suggest connected) issues in current "theory of mind" research:

1. What is involved in grasping the concept of belief, or (to put the same point differently) in having an "explicit understanding" of belief?
2. What explains the striking dissociations that have been found between children's performance on direct and indirect tests for false belief understanding?

In the following two headings we sketch an answer (1). Drawing on this account, in section Understanding the dissociation: theory vs. teleology, we tackle (2).

The concept of belief: reasons vs. laws

We can distinguish two aspects of the causal role of beliefs. One has to do with the input side, the circumstances in which beliefs are acquired and sustained. The other concerns the ways beliefs affect what people do. Someone who has acquired a rudimentary "theory of mind" including the concept of belief—who knows what it is to believe something and who is thus able to have thoughts, beliefs, desires, etc., *about* beliefs—must presumably be familiar, to some extent, with both aspects of the causal role of beliefs. We can put the point by formulating "Introduction" and "Elimination" rules for the concept of belief, comparable (in some ways) to the Introduction and

[2] The adult version of the principle of charity is usually taken to demand that we interpret others in such a way as to make them come out as rational as possible (consistent with the available evidence, of course). See Schueler (2003), Chapter 4, for illuminating discussion.

Elimination rules for the logical constants.[3] The claim would be that mastery of the concept of belief requires being able to reason in accordance with these rules. It would be a difficult task to produce a complete list of the relevant rules, but here an example will suffice:

Introduction rule for Belief

A subject S intentionally puts an object O in location L, is not present when O is subsequently moved elsewhere, and has no reason to think O has been moved.

Therefore, S probably believes that O is still at L.

Elimination rule for Belief

S believes O is at L, and decides to retrieve O.

Therefore S will probably make his/her way to L.

What does it come to, being disposed to follow rules such as these? One influential suggestion is that the thinker must have assimilated a psychological theory, consisting of (more or less platitudinous-sounding) law-like generalizations. On this view, there is a sense in which our disposition to reason in accordance with the Introduction and Eliminations Rules for Belief is underpinned by our possession of a simple theory of belief. For example, our use of the Introduction rule reflects our knowledge that if someone puts an object in a certain place, and does not witness its removal from that place, they tend to believe that the object remains in that place.

We can bring out a basic problem with this account by comparing and contrasting the concept of belief with other psychological concepts. Consider the concept of being drunk. Someone who has acquired the concept of drunkenness is someone who is disposed to draw inferences such as the following:

Elimination rule for Drunkenness

Subject S is drunk.

Therefore, S is probably unsteady on his/her feet.

A salient difference between the two cases is this. Why does drunkenness give rise to its familiar symptoms? Why, for example, does it *impair* rather than *boost* our motor skills? To most of us, the matter is deeply opaque. We have no idea *why* the Elimination rule for Drunkenness holds. No doubt there is a story to be told, tracing the effects of alcohol on the motor system. But you don't need to know that story to know what it is to be drunk: the concept of drunkenness is, in that sense, a relatively shallow concept. The concept of belief differs in this respect. It's not only the experts who understand why the Introduction and Elimination rules for Belief hold. Why should S's belief that O is at L induce him go to L, rather than to dance a jig? The matter is transparent to any reflective thinker who can be said to have a belief-desire psychology (theory of mind): given S's belief and his other circumstances, it *makes sense* for him to go to L—that's where he ought to go (in the "ought of rationality" sense). We have a deeper understanding, in the belief case, of what might be called the rationale of the Introduction and Elimination rules, i.e. the reason why they hold.

[3] To illustrate, the Introduction rule for conjunction is

p

q

———

p & q

For discussion of the relation between understanding and the disposition to reason in accordance with Introduction and Elimination rules (in a range of cases), see Campbell (2002).

Our understanding consists in (a) our ability to reason that S's circumstances (as specified in the Introduction rule) render his belief rational, and that his belief in turn helps to render his action (as specified in the Elimination rule) rational, and (b) our conception of people as rational thinkers and agents. (b) is of course a less austere view than young children's conception of people as responsive to *reasons* (i.e. reason-giving facts). The adult view allows that people may act rationally on the basis of false beliefs and flawed values. And of course we recognize that sometimes people act irrationally.

Belief as perspective: supposition vs. simulation

What's the nature of the reasoning involved in (a)? It is natural, at this point, to turn to the simulation theory. Normally developing humans have a capacity for "imaginative identification" with others. It is in virtue of that capacity, it might be said, that we understand the rationale of the Introduction and Elimination rules. For example, we put ourselves in Mistaken Max's situation, imagine *deciding* to recover the chocolate and *believing* the chocolate to be in the drawer, and then, still within the context of the imaginative exercise, reason to the conclusion "I should go to the drawer." This would be congenial to Jane Heal's and Robert Gordon's views of the role of simulation, which are motivated in part by a concern with the role of rationality in psychological explanation (Gordon, 1995; Heal, 1995).But do we really need to *imagine* having S's mental states to work out that S should go to L? There is a familiar distinction, in the literature on imagination, between supposing and imagining (see, for example, Gendler, 2000; Moran, 1994; Soteriou, 2010). We want to suggest that it is supposition, rather than simulation that holds the key to understanding the rational-explanatory role of beliefs. It's not just that supposition is a more economical procedure than simulation. The important point is that it takes *real* (hypothetical or counterfactual) reasoning, not just imagined or simulated reasoning, to understand what it is rational to do, given the agent's beliefs.

To see the rationale of our Elimination rule for belief it's essential to appreciate that *if* the chocolate were still in the drawer, then this would give Mistaken Max a reason to go to the drawer (i.e. then Max ought—in the "ought of reason" sense—to go to the drawer). To believe that p, after all, is to take it to be a fact that p. And what believing that p makes it rational for one to do depends on what the fact that p would give one a reason to do. Understanding what S's belief makes it rational for him to do thus requires understanding S's *perspective* on what he has reason to do. On the face of it, though, this kind of "perspective taking" is a fairly basic phenomenon—it's not clear that mental simulation necessarily comes into it. The natural way to reach our critical conditional is to reason as follows. "Suppose that the chocolate is still in the drawer. Then what should Max, who urgently needs his chocolate, do? Why, the best course of action, surely, would be for him to go to the drawer." Of course, in one sense, to suppose that p just is to imagine that p. But this is to be distinguished from the richer sense, or senses, of imagination commonly associated with "simulation," such as imagining "from the inside" Max's experiences (e.g. imagining craving chocolate) or the kind of internal play-acting that may be involved in imagining Max's thoughts or propositional attitudes.

Suppositional reasoning involves using as premises propositions one does not believe to be true. But there is nevertheless a sense in which suppositional reasoning is essentially *truth-directed* reasoning. For one thing, we reason from a supposition using the same rules of inference that govern our reasoning from premises we accept. For another, suppositions can be discharged. If you suppose that p, and derive the conclusion that q, you won't of course accept outright that q; but you will, or should, accept outright that *if* p then q. As Dummett puts it: "the point of the procedure [is] that from the fact that certain consequences follow from some hypothesis, we can draw a

conclusion that no longer depends on that hypothesis" (Dummett, 1981, 309). In our example, the *consequence* that follows from the supposition that the chocolate is still in the drawer is that Mistaken Max has a reason to go to the drawer/ought to go to the drawer (in the "ought of reason" sense). The *conclusion* one draws from this is that *if* the chocolate were still in the drawer, Max would have a reason to go to the drawer. This conclusion, in turn, can be used to establish what it is rational for Mistaken Max to do: given that he *believes* the content of our supposition, he ought to go to the drawer (in the "ought of rationality" sense).

One way in which suppositional reasoning differs from simulation is that it is *third-personal.* To determine what someone else would have reason to do under certain suppositions it is not necessary to "recreate" or "replicate" the agent's first personal deliberation. One may think of the agent from a third- (or second-) person perspective: what would he (or you) have reason to do under those suppositions? In contrast, simulating practical reasoning, as standardly conceived, is a matter of imagined, or "make believe," reflection on the question "what should *I* do?" (see Gordon, 1996, 62).

There are a number of considerations to suggest that rational explanation requires suppositional reasoning, rather than imaginative identification. First of all, the idea that it takes an imaginative re-enactment of Max's thought processes to pass a humble false belief task seems suspect on phenomenological grounds. It's not clear, furthermore, why imagination, in the rich sense, should be needed to work out where it makes sense for Max to go, given his belief: straightforward suppositional reasoning seems perfectly adequate to that task.[4] Most importantly, such reasoning would seem to be essential even if, in addition, one performs a practical simulation of Max's deliberation. For imagining someone's reasoning to the conclusion that he should do x does not commit one to the view that it makes sense for him to do x: one can imaginatively re-enact reasoning one takes to be not just based on false premises, but to be confused or deranged. Insofar as a practical simulation is to enable one to appreciate the rationality of Max's action, it has to reflect one's independent judgment that, given Max's belief, it's rational for him to go to the drawer. That judgment cannot itself be based on simulation. It requires reasoning to the conclusion that, if the chocolate were in the drawer, there would be a justifying reason in favor of Max's going there; and that therefore, given his belief that the chocolate *is* in the drawer, he ought to go there (in the "ought of rationality" sense). There are two features of such reasoning that bear emphasis. One is that it embeds the simple kind of teleological reasoning at which (we argued) even young children are quite proficient. The other feature is that it requires the reasoner to reason (to *genuinely* reason—not just to *pretend* reasoning) from premises she regards as false, in order to derive true conclusions concerning what's rational for others to do. We call such reasoning "teleology-in-perspective," to highlight both its continuity with young children's simple teleology and the fact that it presents its practitioners with a *perspective problem*. They need to be able to move back and forth between two conflicting points of view on what someone has reason to do.

To sum up our discussion so far: young children's performance on classical false belief tests reflects both a vital insight—people generally do what it makes sense for them to do—and a crucial limitation—their inability to understand that it can be rational for someone to do something even if there is no objective reason for them to do it. This limitation disables young children from fully grasping the concept of belief: they are unable to understand why believing something has the causal role it does, i.e. to recognize what we called the "rationale" for the Introduction and Elimination rules. The next question is this: how might this account help to shed light on the dissociation between children's performance on direct and indirect tests?

[4] See Millar (2004) for illuminating discussion of this point.

Understanding the dissociation: theory vs. teleology

Children's understanding of the role of belief in intentional action has been intensively investigated with the "Mistaken Max" false belief task (Wimmer & Perner, 1983). When Max returns looking for his chocolate, 3-year-old children answer with the actual location (cupboard), while 5-year-olds answer with the location Max believes the chocolate to be in (drawer). Many studies tried to find ways of demonstrating earlier understanding, but a large meta-analysis (Wellman, Cross, & Watson, 2001) of these studies showed that the understanding that action depends on belief develops around 4 years of age.

A dissociation

Clements & Perner (1994) found a dissociation between different measures of understanding. The paradigm was slightly changed. Sam the Mouse used different exits from his abode when looking in one or the other of two boxes outside. He had put a piece of cheese in one box (box 1), then went inside to sleep. While asleep someone transferred his cheese to the other box (box 2). This set up allowed the filming of the children's eye gaze when Sam woke up with a craving for his cheese. Most 3-year-olds looked for Sam in expectation of his reappearance at the exit to box 1 (where he thought his cheese was). This occurred only in the false belief condition, but not in a true-belief control condition, where Sam had seen the transfer to box 2. Most interestingly, all the young 3-year-olds who showed this looking behavior still maintained, when asked, that Sam would come out from the exit to box 2 (where the cheese actually was).

This dissociation has been replicated (Garnham & Perner, 2001) by different investigators (Low 2010; Ruffman, Garnham, Import, & Connolly, 2001; Wang, Low, Jing, & Qinghua, 2012). Perner & Clements 2000 made a case that children's anticipatory looking shows the characteristics of indirect measures indicative of implicit knowledge (Reingold & Merikle 1988), e.g. guessing by blindsight patients of the location of a stimulus in their blind field (Weiskrantz, 1986), by sighted persons in the Roelof's induced motion illusion (BridgemanKirch, & Sperling, 1981; Bridgeman, Peery, & Anand, 1997), thumb-finger span size indicating an object's true size when explicit size judgments are distorted by illusion effects (Agliotti, DeSouza, & Goodale, 1995; Stöttinger, Aigner, Hanstein, & Perner, 2009; Stöttinger, Soder, Pfusterschmied, Wagner, & Perner, 2010).[5] Furthermore, Ruffman et al. (2001) showed that children seem absolutely unaware—don't even

[5] The distinction between direct and indirect tests is not as obvious as it may seem. Naively one would think that a direct false belief test is one in which the child is asked directly about an agent's belief. In that case, the good old standard false belief test would be indirect, because children are not asked about Mistaken Max's belief, but about his future action. Hence, the test should strictly speaking not be called a direct false belief test, but a direct test of mistaken intentional action.

In general other problematic aspects are that the question may be directly about the matter of interest but still count as indirect. For instance, when a blindsight person is asked to guess where a stimulus is, implicit knowledge can be used, but not when asked to point to where the stimulus actually is. The same can be shown with normally sighted persons when they have to indicate a near threshold change of brightness (Marcel, 1993). So the critical feature is not the form of the question but how the question is to be taken. If the respondent is to take it as a request to say where something really is, then it is a direct test. If the question is to be taken as where something could be (a blind guess), then it is an indirect test, because the pointing gesture to where it could be (guess) is influenced by where it actually is. Moreover, when blindsight patients are asked to insert their hand into a slot of different orientation they can do so above chance even when they can't consciously see the slot, but they cannot indicate with their hand the direction of the slot (Perenin & Rosetti, 1996). Similar abilities have been reported with healthy persons' susceptibility to illusions.

consider it a vague possibility—that the agent could reappear where they look in anticipation to see him reappear.

Early sensitivity—the facts

The use of indirect tests has led to the discovery of very early sensitivity to agent's false beliefs. We can distinguish four different paradigms.

Looking in expectation

Children look in expectation where they expect a hand to appear on the basis of where the agent thinks an object is. This can be shown by 2 years (Southgate, Senju, & Csibra, 2007) and perhaps earlier (Neumann, 2009; Southgate, 2008).

Looking time

Infants of about 14 months look longer at the test scene when a mistaken agent searches in the correct location than when she searches in the wrong location where she thinks the object is (Onishi & Baillargeon, 2005). The longer looking is interpreted as infants expecting a different action (search in the empty container where the agent believes the object to be) than what is shown (agent searches in correct container). Hence this method has been dubbed "violation of expectation paradigm." This finding has led to an explosion of demonstrations that infants in their second year expect agents to act according to their beliefs and not the real state of affairs.

A somewhat different use of looking time differences was made by Kovacs, Teglas, & Endress (2010). As early as 7 months, infants' looking time was recorded when discovering a surprising outcome—a ball behind a screen had disappeared. Their looking time was longer when a bystander shared their belief that the ball was still behind the screen than when the bystander thought the ball had disappeared. A similar technique using reaction times has been pioneered by Apperly, Riggs, Simpson, Chiavarino, & Samson (2006) assessing automaticity of belief attribution in adults with the conclusion that it is not automatic.

Interpretation of referential expressions

This paradigm was pioneered by Carpenter, Call & Tomasello (2002) and Happé & Loth (2002) with children around the age of 3 years. Southgate, Chevallier, & Csibra(2010) tested 17-month-old infants who watched an agent place two novel, unnamed objects in two separate boxes. Unbeknownst to the agent, the contents were then switched. When the agent returned, she pointed to a box (the incorrect box in the false-belief condition) and said: "Do you remember what I put in here? There's a *sefo* in here. There's a *sefo* in this box. Shall we play with the *sefo*?" In the false-belief conditions, children correctly chose the item in the other box, not the one the experimenter pointed to. The authors' interpretation is that children understood that the experimenter wanted from the indicated box the object she thought was in there and not the one that was actually in there.

Helping behavior

In the false-belief condition by Buttelmann, Carpenter, and Tomasello (2009), an agent failed to witness her favourite toy being moved and returned to the first box to retrieve the toy, but couldn't get to open the box. Children were then asked to "help" the agent. Over 70% of 18-month-olds approached the second box. In contrast, less than 20% did so in a knowledge condition where the unsuccessful agent had witnessed the transfer to the new box but tried to open the empty box. Buttelmann et al. (2009) suggested that toddlers approached the second box in the false-belief condition because they recognized that the agent falsely believed that her toy was still inside the first box and concluded from the agent's unsuccessful attempt to open that box that she wanted to

retrieve the toy she thought was in that box but which was now in the new box. So the child had to orient to the new box to retrieve the desired toy.

Early sensitivity—interpretation

One question about these findings concerns the best way to characterize the two groups of tasks—those that reveal early sensitivity and those traditional tasks that point to later understanding. Clements and Perner (1994; Perner & Clements, 2000) characterized the tasks as indirect and direct, inspired by the use of this terminology in the consciousness literature (Reingold & Merikle, 1988). More recently, Scott and Baillargeon (2009) characterized the difference as one of "spontaneous" and "elicited" responses, which has much to recommend itself, but also does not quite capture the relevant difference, as the authors themselves imply (p. 391): "Finally, infants and toddlers should succeed at indirect-elicited-response tasks that require them to respond to questions or prompts that only indirectly tap their representation of an agent's false belief." And the authors refer to the studies by Buttelmann et al (2009) and Southgate et al (2010) as good examples. So "indirectness" seems the critical factor.

Several distinctions have been proposed to characterize the difference between the kinds of knowledge underlying the early sensitivity and later understanding:

1. Implicit—explicit (Clements & Perner, 1994; Perner & Clements, 2000)
   (a) unconscious—conscious (Garnham & Perner, 2001; Ruffman et al., 2001)
   (b) procedural—declarative
   (c) non-conceptual—conceptual (Rakoczy, 2012)
   (d) automatic (spontaneous)—controlled (Apperly et al., 2006)
2. Modular—central process (Leslie, 1994)
3. Causal understanding: shallow—deep (behavior rules—mental state rules; Perner, 2010)

But from these characterizations no general principles follow for a detailed account, which would answer the following two questions:

1. Why does early sensitivity emerge only in indirect tasks and not the traditional direct ones?
2. Why do the younger children give systematically wrong answers on the direct tests until they are about 4 years old?

One detailed account was provided by Scott and Baillargeon (2009). They propose two subsystems to the theory of mind system SS1 and SS2 (pp. 1174–5):

> SS1 allows them to attribute two kinds of internal states to the agent: motivational and reality-congruent informational states … Motivational states specify the agent's motivation in the scene and include goals and dispositions. Reality-congruent informational states specify what knowledge or accurate information as construed by the infant the agent possesses about the scene… SS2 extends SS1 in that it allows infants to attribute reality-incongruent informational states to agents. When an agent holds a false or a pretend belief about a scene …

Scott and Baillargeon's answer to our two questions rests on the assumption that indirect tests require only representation of belief, while direct tests require the interplay of three processes (p. 1176):

> We assume that success in the Sally–Ann task depends on the interaction of three separate processes. First, children must represent Sally's false belief about the marble's location; … Second, when asked the

> test question, children must attend to the question, decide to answer it, and tap their representation of Sally's false belief response-selection process. Finally, children must inhibit any prepotent tendency to answer the question based on their own knowledge of the marble's current location response inhibition process. Children then fail because (a) the joint activation of the false-belief-representation process and the response-selection process overwhelms their limited information-processing resources, and/or (b) the neural connections between the brain regions that serve these two processes are still immature and inefficient in early childhood.

This answers question 1: children show sensitivity to false belief in indirect tests earlier than in direct tests because indirect tests tax their limited processing system less than direct tests. An answer to question 2 is not obvious. If the overload leaves the toddler without any means to answer then the child can but guess, but not be systematically wrong. If the overload disables SS2, but leaves SS1 as default, the consequence would be that SS1 would represent the agent's ignorance of the new location and the child, again, can but guess what the agent will do.

We can see two interesting weaknesses in this approach. The one, already discussed, is the need to explain the systematic errors on the direct tests. The other weakness is that it does not explain why only direct tests require response selection and inhibition but not also indirect tests. Scott, Baillargeon, Song, & Leslie (2010, p. 391) have this to say:

> In marked contrast, success in spontaneous-response tasks such as VOE [violation of expectation] and AL [anticipatory looking] tasks depends on only one process, the false-belief-representation process; the response-selection and response-inhibition processes are not activated because children produce their responses spontaneously rather than in answer to direct questions.

So the response is given spontaneously without external prompt by a question,[6] but that still leaves the child to select one of many responses (e.g. should I look to location 1 or to location 2 if I want to see him come out of the exit). Selection of the correct looking response would also be interfered with by the tendency to look to the exit near the desired object's real location, which needs to be inhibited. In fact, if selection of the spontaneous looking response tends to be automatic and implicit in contrast to an explicit response to a question, then it should, if anything, be more difficult to inhibit the automatic response than the more explicit response—exactly the opposite of what is being observed.

From causes of behavior to reasons for acting as causes

Our proposal also assumes two different approaches (or systems). One consists of a purely nomic causal understanding of *behavior* as caused by motivational and informational states in the tradition of theory theory (Gopnik and Meltzoff, 1997; Leslie, 1994). It underlies the data based on indirect tests. The other consists of understanding reasons for *action* in the tradition of those who emphasize the role of rationality (and the principle of charity) in interpretation (Davidson, 1963; McDowell, 1985). It is triggered by direct tests and is based on expecting people to act in a way they have reasons to act.

Approach 1: caused behavior

For present purposes we'd like to remain completely agnostic about the causal depth of this understanding, e.g. whether infants make causally shallow connections from observable indicators of

[6] In fact the prompt need not be a question. Garnham and Perner (2001) had children place a mat to catch the returning agent. Children who placed the mat spontaneously without hesitation tended to place the mat to the exit where the agent thought his object was, while children who needed prompting in the form of a reminder to move the mat tended to place it at the exit where the object really was.

motivational and informational states to future behavior (behavior rules, Povinelli & Vonk, 2004) or whether they infer inner states from these indicators, and predict and interpret behavior on the basis of these inner states (e.g. Tomasello, Call, & Hare, 2003).[7] We also stay neutral about origin. Infants' knowledge might be innate (and modular) and emerge at particular times by maturation (Leslie, 1994), or be rapidly built up by statistical learning (Ruffman, Taumoepeau, & Perkins, 2012) or be acquired through a theorizing process (Gopnik & Meltzoff, 1997).

We have commitment—though not irrevocable—on some of the other features of its knowledge base. Evidence suggests that it is based on implicit knowledge. The dissociation observed by Clements and Perner (1994) and children's reluctance to acknowledge the agent to reappear where they looked in anticipation (Ruffman et al., 2001) suggests that it is not consciously accessible. Consequently, it is likely to be automatic and not under voluntary control (Apperly & Butterfill, 2009). Moreover, indirect measures tend to consist of online "reactions to the unfolding events" (Scott and Baillargeon, 2009, p. 1176) or live interactions in the helping paradigms. This also suggests that the knowledge is procedural and may not be available for conditional reflection, which requires declarative knowledge.

The general characterization of this approach is that it treats behavior of organisms or moving dots on a par with the movement and changes of inanimate physical objects. Theory of mind is just one theory among many others.

Approach 2: reasons for action as causes

To appreciate behavior (goal-directed movement) as intentional action, one has to understand it as behavior for which the agent has reasons. If one were in Mistaken Max's situation, having put one's chocolate earlier into the drawer and now looking for it, one does not ask oneself what one *will* do next, but what one *should* do next—go to the drawer or the cupboard? Since from one's own point of view, the chocolate is still in the drawer, one seems to have good (objective) reasons to go to the drawer. And because one is motivated by this fact, one knows that one is likely to go there because one should go there. One does not simply conclude this on the basis of a law-like regularity: "Whenever I want something and think (know) that it is in location x then I will go to location x."

To understand others as intentional agents is to understand what they are doing or will do in terms of what they should do given their goals and circumstances.[8] Rakoczy, Warnecken, and Tomasello (2008) showed that children as young as 2 years expect other people to act as they should. If the pronounced goal is to play a certain game then one should behave according to the game's constituent rules. Or else 2- and particularly 3-year-olds get very upset. Children's normative attitude is based on understanding objective reasons, i.e. teleology (Perner & Roessler, 2010). Teleology captures intentional actions very well, as long as they are based on objective goals and objectively appropriate instrumental actions.[9]

[7] Despite the recent evidence of early understanding of the mind the critical evidence whether this understanding is based on behavior rules or mental state computation is still outstanding (Perner, 2010).

[8] We want to emphasize (see Belief as perspective: supposition vs. simulation) that this should not be understood in the sense of simulation theory as imaginatively putting oneself into the other person's situation. It only requires seeing from one's own position what is needed (goal) and what needs to be done to achieve it (instrumental action) by whoever is in a position to carry out that action.

[9] This bears resemblance to Scott and Baillargeon's (2009) subsystem SS1 as it involves goals and—to use their term—reality-congruent instrumental actions.

Teleology breaks down as a means of understanding intentional actions when subjective perspectives on goals and means are involved. Mistaken Max's move to the empty drawer remains an irrational behavior for the young teleologist. The teleologist can, however, recapture Mistaken Max's rationality by realizing that he is mistaken, i.e. has a deviant perspective on the world, and employ teleology within his perspective (Perner, 2004). Earlier we labelled this kind of reasoning "teleology in perspective" (see Belief as perspective: supposition vs. simulation; see also Perner & Roessler, 2010). Importantly, teleology-in-perspective preserves the rationality of Mistaken Max's action: he can be seen to act on the basis of what from his perspective appears to be an objective reason. This ability becomes operative around 4 years when children develop some notion that different perspectives exist and, thus, need not anymore rely on being switched to another person's perspective but can voluntarily do so (Perner, Stummer, Sprung, & Doherty, 2002).

There is now a large amount of evidence that children at this age become able to succeed on a large variety of otherwise unrelated tasks that share the need for perspective understanding. For instance, level 2 perspective taking (Masangkay, McCluskey, McIntyre, Sims-Knight, Vaughn, & Flavell, 1974), interpreting ambiguous drawings (Doherty & Wimmer, 2005), understanding false direction signs (Leekam, Perner, Healey, & Sewell, 2008) alternative naming (Doherty & Perner, 1998; Parkin, 1994; Perner et al., 2002), episodic memory (Perner & Ruffman, 1995; Perner, Kloo, & Stöttinger, 2007; Sabbagh, Moses, & Shiverick, 2006), and understanding identity information (Perner et al., 2010) not only emerge at this age, but also correlate specifically with the traditional false belief task.

A main purpose of understanding reasons for action is to explain, rather than predict behavior (Andrews, 2012) and be able to reason and argue about the correctness or appropriateness of one's own and others' conduct. It is an essential glue of human society and tied to linguistic interaction. Its knowledge base must be explicit: declarative (non-procedural) to be used in conditional arguments, access conscious, conceptual for linguistic exchanges, and under voluntary control (at least for voluntary retrieval). It cannot be modular, since it needs to be accessible for argumentation.

Explaining the evidence: answering our two questions

Having described the two approaches taken by children (also by adults) we need to check how well this proposal can answer our two questions (from section Early sensitivity—interpretation):

1. Children show sensitivity to false beliefs very early on indirect measures in online and interactive tasks because they have implicit nomic knowledge about motivational and informational states causing behavior (we are non-committal as to causal depth of this knowledge). In contrast, on direct tests the knowledge in question is part of the test specification ("Where will Max go?") which requires a declarative commitment, which—as the consciousness literature suggests—requires *explicit knowledge*. For this the young children employ pure teleology; they make predictions of what someone will do in terms of what the person should do, i.e. has objective reasons to do.
2. Children's predictions on direct false belief tests show the reality error because they are teleologists. Around 4 years they become aware of the existence of perspective differences, which enables them to see a person's reasons relative to a different perspective (teleology in perspective). The age point conforms to the age at which many other tasks that require awareness of perspectives are mastered.

Conclusion

We drew attention to a feature of how we understand intentional action that tends to get lost in the theory theory of mind. Our naïve belief-desire psychology is not primarily a body of law-like generalizations of how agents tend to behave, but involves an understanding that they act for reasons. They do what they should do—for the most part. We ventured the contention that infants' and toddlers' early expectations of how people will act, especially when a false belief is involved, may be based on law-like generalizations, which remain implicit and dissociate from an understanding of people acting for reasons until they become able to understand reasons relative to an agent's perspective around 4 years. Because of the dissociation, we think that the earlier understanding is implicit and the later understanding explicit. Our mentalist understanding (theory of mind) of intentional actions has to become explicit at some point anyway, since one of its prime functions is to argue about and justify conduct. Although we emphasized the difference in knowledge base of earlier and later understanding, we also like to think that there is developmental continuity between the approaches (this is one reason we do not want to talk of systems that are often associated as independent); there is evidence that performance in direct tasks is related to earlier performance on indirect tasks (Low, 2010; Thoermer, Sodian, Vuori, Perst, & Kristen, 2011). In particular, the discrepancy between how mistaken people act and the young teleologist's wayward predictions must be an important motor for moving to a more sophisticated understanding. As Karmiloff-Smith and Inhelder (1974) observed in the context of children's understanding of how to balance objects on a fulcrum: "If you want to get ahead, get a theory." So, in our case, infants have an implicit sense of how people under certain informational conditions are likely to act, then they get a rough theory (teleology) that people act as they should act, which they then need to refine into teleology in perspective.

References

Agliotti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, **5**(6), 679–85.

Andrews, K. (2012). *Do Apes Read Minds: Toward a New Folk Psychology.* Cambridge: MIT.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116**, 953–70.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, **17**, 841–4.

Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics*, **29**, 336–42.

Bridgeman, B., Peery, S., & Anand, S. (1997). Interaction of cognitive and sensorimotor maps of visual space. *Perception and Psychophysics*, **59**(3), 456–69.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, **112**(2), 337–42.

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

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology*, **20**, 393–420.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, **9**, 377–97.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, **60**, 685–700.

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

Doherty, M. J., & Wimmer, M. (2005). Children's understanding of ambiguous figures: Which cognitive developments are necessary to experience reversal? *Cognitive Development*, **20**, 407–21.

Dummett, M. (1981). *Frege. Philosophy of Language,* 2nd edn. London: Duckworth.

Garnham, W. A., & Perner, J. (2001). When actions really do speak louder than words—but only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology*, **19**, 413–32.

Gendler, T. S. (2000). The puzzle of imaginative resistance. *Journal of Philosophy*, **97**, 55–80.

Gopnik, A., & Meltzoff, A. N. (1997). *Word, Thoughts, and Theories*. Cambridge: Bradford Books, MIT Press.

Gordon, R. (1995). Folk psychology as simulation. In M. Davies, & T. Stone (Eds), *Folk Psychology*. Oxford: Blackwell.

Gordon, R. M. (1996). "Radical" simulationism. In P. Carruthers, & P. K. Smith (Eds), *Theories of Theories of Mind,* pp. 11–21. Cambridge: Cambridge University Press.

Happé, F., & Loth, E. (2002). "Theory of mind" and tracking speakers' intentions. *Mind and Language*, **17**, 24–36.

Heal, J. (1995). "Replication and Functionalism'. In M. Davies & T. Stone (Eds), *Folk Psychology*. Oxford: Blackwell.

AQ1: Please cite as Hornsby ref in text

Hornsby, J. (2008). A disjunctive conception of acting for reasons. In A. Haddock, & F. Macpherson (Eds), *Disjunctivism*. Oxford: Oxford University Press .

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, **3**, 195–212.

Kolodny, N. (2005). Why be rational? *Mind*, **114**, 509–63.

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science New York*, *NY*, **330**(6012), 1830–4.

Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal of Developmental Psychology*, **26**(4), 485–97.

Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, **50**, 211–38.

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, **81**(2), 597–615.

Marcel, A. J. (1993). Slippage in the unity of consciousness. In G. R. Bock, & J. Marsh (Eds), *Experimental and Theoretical Studies of Consciousness* (pp. 168–86). Chichester: Wiley.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, **45**, 357–66.

McDowell, J. (1985). Functionalism and anomalous monism. In E. LePore, & B. McLaughlin (Eds), *Actions and Events*. Oxford: Blackwell.

Millar, A. (2004). *Understanding People*. Oxford: Clarendon Press.

Moran, R. (1994). The expression of feeling in imagination. *Philosophical Review*, **103**, 75–106.

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

Neumann, A. (2009). Infants' implicit and explicit knowledge of mental states. *Evidence from Eye-tracking Studies*, Chapter 7. Ketsch Verlag: Microfiche.

Parkin, L. J. (1994). *Children's understanding of misrepresentation.* Unpublished Thesis for doctorate, University of Sussex.

Perenin, M. T., & Rossetti, Y. (1996). Grasping without form discrimination in an hemianopic field. *Neuroreport*, **7**, 793–7.

Perner, J. (2004). Wann verstehen Kinder Handlungen als rational? In H. Schmidinger & C. Sedmak (Eds), *Der Mensch—ein"animal rationale"? Vernunft—Kognition—Intelligenz* (pp. 198–215). Darmstadt: Wissenschaftliche Buchgesellschaft.

Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P. A. Frensch, & R. Schwarzer (Eds), *Cognition and Neuropsychology: International Perspectives on Psychological Science,* Volume 1 (pp. 241–61). London: Psychology Press.

Perner, J., & Clements, W. A. (2000). From an implicit to an explicit theory of mind. In Y. Rossetti, & A. Revonsuo (Eds), *Beyond Dissociations: Interaction Between Dissociated Implicit and Explicit Processing* (pp. 273–93). Amsterdam: John Benjamins.

Perner, J., Kloo, D., & Stöttinger, E. (2007). Introspection and remembering. *Synthese*, **159**, 253–70.

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science*, **333**, 474–7.

Perner, J., & Roessler, J. (2010). Teleology and causal reasoning in children's theory of mind. In J. Aguilar, & A. Buckareff (Eds), *Causing Human Action: New perspectives on the causal theory of action* (pp. 199–228). Cambridge: Bradford Books, MIT Press.

Perner, J., & Ruffman, T. (1995). Episodic memory an autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. Special Issue: Early memory. *Journal of Experimental Child Psychology*, **59**(3), 516–48.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. Cognitive Development, Special Issue on "Constructivism Today" as part of the inauguration of Cognitive Development as the official journal of the Jean Piaget Society for the Study of Knowledge and Development. *Cognitive Development*, **17**, 1451–72.

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

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology*, **30**, 59–74.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental Psychology*, **44**, 875–81.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception and Psychophysics*, **44**(6), 563–75.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, **80**, 201–24.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, **30**, 87–104.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs and false signs. *Child Development*, **77**, 1034–49.

Schueler, F. (2003). *Reasons and Purposes*. Oxford: Clarendon.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, **80**(4), 1172–96.

Scott, R. M., Baillargeon, R., Song, H. J., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, **61**(4), 366–95.

Soteriou, M. (2010). Cartesian reflections on the autonomy of the mental. In J. Aguilar, A. Buckareff & K. Frankish (Eds), *New waves in philosophy of action* (pp. 122–40). Palgrave: Macmillan.

Southgate, V. (2008). "Attributions of false belief in infancy." Invited presentation at the EPS Research Workshop on Theory of Mind: A workshop in celebration of the 30th anniversary of Premack & Woodruff's seminal paper, "Does the chimpanzee have a Theory of Mind? (BBS 1978), University of Nottingham, 11–12 September

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, **13**(6), 907–12.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, **18**, 586–92.

Stöttinger, E., Aigner, S., Hanstein, K., & Perner, J. (2009). Grasping the diagonal: controlling attention to illusory stimuli for action and perception. *Consciousness and Cognition*, **18**(1), 223–8.

Stöttinger, E., Soder, K., Pfusterschmied, J., Wagner, H., & Perner, J. (2010). Division of labour within the visual system—fact or fiction?—Which kind of evidence is appropriate to clarify this debate? *Experimental Brain Research*, **202**, 79–88.

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2011). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology*, **30**, 172–87.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees versus humans: It's not that simple. *Trends in Cognitive Sciences*, 7, 239–40.

Wang, B., Low, J., Jing, Z., & Qinghua, Q. (2012). Chinese preschoolers' implicit and explicit false-belief understanding. *British Journal of Developmental Psychology*, **30**, 123–140.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications.* Paperback edn 1998. Oxford: Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, **72**, 655–84.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**, 103–28.

Woodward, J. (2011). Causal perception and causal cognition. In J. Roessler, H. Lerman & N. Eilan (Eds), *Perception, Causation, and Objectivity* (pp. 229–63). Oxford: Oxford University Press.