Representing beliefs: A matter of perspective awareness and counterfactual reasoning?

Description of the symposium topic.

Children take a very important step towards social understanding when they are able to represent a belief. There is an ongoing discussion about whether this ability is an early emerging, modular process or whether it is a gradually evolving set of cognitive processes which may result from an interaction between genetics and environment (Ruffman & Perner, 2005). For example, it has been shown that one year old infants can correctly anticipate the action of an agent who has a false belief about the location of an object (Onishi & Baillargeon, 2005; Southgate et al., 2007). On the other hand, performance on the classical false belief task (Wimmer & Perner, 1983) lies below chance when children are 41 months old, and reaches above chance when children turn 48 months (Wellman et al., 2001). The present symposium will cover different positions in this discussion. The first two talks in the symposium will address the question of which cognitive processes underpin the ability to represent a belief based on evidence from developmental and functional neuroimaging research. The third talk highlights an apparent inconsistency between the claim made by the first two talks and the claim that infants can represent beliefs from around their first birthday and attempts to resolve the apparent inconsistency by considering the possibility that there are different ways of representing a belief. Taken together, the three talks show that being able to represent beliefs involves being able process perspective differences and reason counterfactually, and that (properly understood) these findings are not only consistent with evidence of mindreading in infants but may also illuminate the nature of infants' competence.

The first talk (Matthias Schurz) tries to distinguish different cognitive processes that contribute to theory of mind reasoning based on findings from functional neuroimaging. Brain

regions engaged by different types of activation tasks used to study theory of mind are meta-analyzed and compared. An attempt to delineate cognitive components of theory of mind reasoning is made by dividing activation tasks in those which require processing of a perspective difference and those which do not. By perspective difference, the author refers to a difference in content between different representations of the same representational target, for example an object, scene or an event (Perner et al., 2002). A well known example of a theory of mind task which creates a perspective difference is the false belief task, where a protagonist's belief contrasts with the participant's own view of reality. Popular examples of theory of mind tasks which do not create perspective differences are tasks which ask participants for a mental-state judgement based on the picture of a human face or tasks which present moving triangles (participants view animations of simple geometrical shapes moving in a way that implies intentional actions). Results of the meta-analyses show areas of activation common to theory of mind tasks with and without perspective, and also areas only activated by one or the other type of task. Interestingly, the brain areas selectively engaged by theory of mind tasks which present a perspective difference are also engaged by processing of perspective differences in other domains. For example, the same brain area shows increased activation for "remember" judgements (which require understanding that the re-experience of a former event provides a perspective of the originally experienced event) compared to "know" judgements in episodic memory tasks. The talk also discusses whether understanding perspective differences is a prerequisite for understanding false belief.

The second talk (Eva Rafetseder) more specifically addresses the question of whether counterfactual reasoning is a prerequisite for reasoning with false beliefs. That counterfactual reasoning is a necessary condition is suggested in a developmental study by Riggs et al. (1998). Performance on the false belief question ("Where will Maxi look for his chocolate?") and on the counterfactual question ("If mother had not baked a cake, where would the chocolate be?") was highly correlated. However recent findings indicate that children's

correct answers to counterfactual questions are not always based on counterfactual reasoning (CFR) but sometimes involve using only basic conditional reasoning (BCR: Rafetseder, Cristi-Vargas, & Perner, 2010), i.e., applying universally quantified conditionals that express general regularities. Use of CFR tends to emerge not before the age of 6 years. Two studies investigated how children (7-14 years) perform on false-belief tasks when controlling for answers based on BCR. Both studies found a highly significant correlation ($r = .58$) between children using CFR to answer the counterfactual question and their correct answers to the false belief question. Moreover, hardly any children gave correct answers to the belief and wrong answers to the counterfactual question, suggesting that counterfactual reasoning is a prerequisite for predicting actions based on false beliefs.

The third talk (Stephen Butterfill) takes the findings of the first two talks as a starting point. According to these findings, being able to represent false beliefs involves being able to (i) process perspective differences or (ii) reason counterfactually (or both). However, infants around their first birthday cannot process perspective differences nor reason counterfactually; but they can pass some false belief tasks (Onishi & Baillargeon, 2005; Southgate et al., 2007). It is argued that this discrepancy cannot be explained merely by distinguishing between implicit and explicit representations of false beliefs in part because one-year-old infants also succeed on false belief tasks which involve actively helping others, interpreting their utterances and pointing to provide information (Buttelman et al., 2009; Knudsen & Liszkowski, 2011; Southgate et al., 2010) and it would be natural to suppose that success on these sorts of tasks reflects explicit representations of false belief.

In order to explain this discrepancy it may be necessary to take into account that representations of beliefs can have different levels of complexity. An analogous example for differences in the complexity of representations is provided: In order to represent the efficiency of a car, one can merely use the driven distance in relation to fuel consumption.

However, one could also use a more accurate measure which relates force to fuel consumption. Similarly, representations of beliefs can vary in terms of their complexity, and different false belief tasks may require different levels of representational elaboration. In addition, it is suggested that the early mastery of some false belief tasks involves only one representational system while the late mastery of the other false belief tasks involves multiple representational systems. This can be seen in analogy to children's ability to understand numerosity: infants have a vague sense of quantity (e.g., they can discriminate between small and large sets of objects) already before they learn to count (by which they become able to precisely discriminate between large numerosities).

Based on the latter two arguments, the third talk offers an integrative framework on the cognitive underpinnings of belief representation: Infants from one year of age on *can* represent false beliefs using *a comparatively simple measure* in a *modular process*. However, infants at that age *cannot* represent false beliefs using *a comparatively sophisticated measure* in a *non-modular processes*. This additionally requires the ability to (i) process perspective differences or (ii) reason counterfactually (or both), as the first two talks demonstrate.

References

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

Knudsen, B. & Liszkowski, U. (forthcoming 2011). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*.

Onishi, K. H, & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719),255–258.

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, 17*, 1451–72.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development, 81*(1), 376–389.

Riggs, K. J., Peterson, D. M., Robinson, E. J. & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13*(1), 73–90.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences, 9*(10), 462–463.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. Developmental Science, 13(6), 907–912.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587–592.

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

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128.

Neural bases of perspective differences in theory of mind tasks: A meta-analysis of neuroimaging studies.

Matthias Schurz, University of Salzburg

Functional neuroimaging research suggests that different neuro-cognitive processes are contributing to theory of mind reasoning (e.g., Saxe & Young, in press). Studies from our group suggested that, in addition to the processing of mental states, theory of mind reasoning sometimes also engages the processing of a perspective difference, which is related to activation in the left temporo-parietal junction (e.g., Aichhorn et al., 2006, Perner et al., 2006). A well-known example for a task which engages processing of perspective differences is the false belief task, which requires distinguishing between the false belief of a protagonist and one's own view of reality. Based on a graphical summary of imaging studies, Perner & Leekam (2008) observed that theory of mind tasks that present a perspective difference (e.g. understanding false beliefs of others) activate a different portion of the left temporo-parietal junction compared to tasks that do not present a perspective difference (e.g. assess mental dispositions of others). In contrast, both groups of tasks activated roughly the same portion of the right temporo-parietal junction. In the present study, we performed a meta-analysis on brain imaging findings to provide a quantitative evaluation of the observations of Perner & Leekam (2008). In addition, we compared brain activation for theory of mind tasks to brain activation for other groups of tasks which may involve processing of perspective differences, in particular episodic memory or mental rotation tasks.

We conducted a quantitative, effect-size based meta-analysis using the SDM software (http://sdmproject.com, Radua et al., 2011). For each group of studies, we calculated a meta-analytic map at an uncorrected threshold of $p < .001$ and a minimum cluster size of 10 voxels. In a first step, we performed a meta-analysis on neuroimaging studies on theory of mind

reasoning. We separated studies into those presenting theory of mind tasks which create a perspective difference and those presenting theory of mind tasks which do not create a perspective difference. A statistical comparison between the meta-analytic maps for the two groups of studies was performed by calculating a linear contrast (see Radua et al., 2011). In a second step, we performed meta-analyses of neuroimaging studies on episodic memory and mental rotation. Areas commonly activated by processing of perspective differences in theory of mind tasks and episodic memory or mental rotation tasks were determined by conjunction analyses in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/).

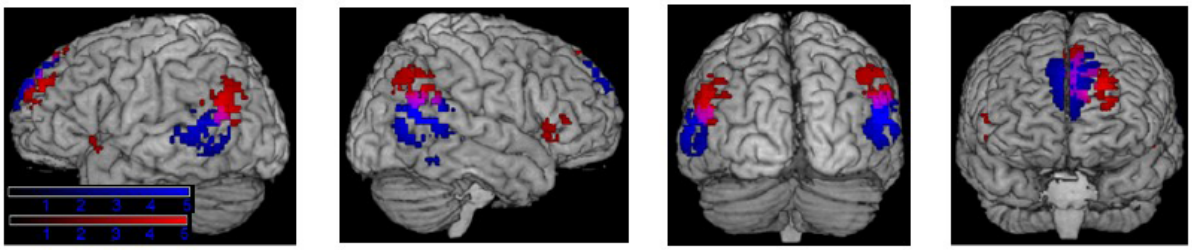
Consistent with recent reviews (e.g., Mar, 2011; Saxe & Young, in press), our meta-analysis found the medial prefrontal cortex and the bilateral temporo-parietal cortices reliably activated in theory of mind tasks. Tasks that did not require processing of a perspective difference predominantly activated ventral/anterior aspects of the temporo-parietal cortices, in particular the middle temporal gyri and the posterior superior temporal sulcus (see Figure 1A). In contrast, tasks that did require processing of a perspective difference (Figure 1B) activated more dorsal/posterior aspects of the temporo-parietal cortices, in particular the angular gyrus. A statistical comparison between the meta-analytic maps for tasks with and without perspective differences confirmed this task-dependent dissociation in terms of brain activation both for the left and the right temporo-parietal cortex (Figure 1C). A conjunction analysis between activation for theory of mind tasks presenting a perspective difference and activation for episodic memory tasks found an area of overlapping activation in the left temporo-parietal junction, with an activation peak in the left angular gyrus. No overlap was found for the right temporo-parietal junction. Results of the conjunction analysis between activation for theory of mind tasks presenting a perspective difference and activation for mental rotation tasks were less clear cut.

Results from our meta-analysis are consistent with the idea that more dorsal aspects of the left temporo-parietal junction, in particular the left angular gyrus, are engaged in the

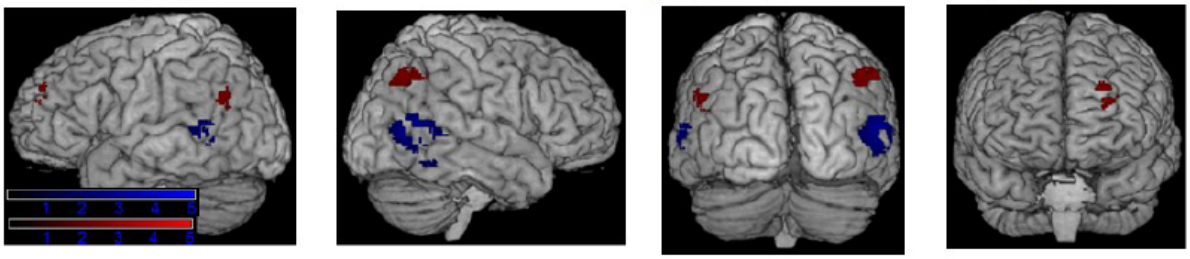
processing of a perspective difference in theory of mind tasks. The overlap in activation with episodic memory tasks reinforces the idea of Aichhorn et al. (2006) and Perner et al. (2006) that processing of perspective differences is not a process specific for theory of mind reasoning, but a common element of tasks that require sensitivity to perspective. For example, it was argued that the remember/know decision often used in episodic memory tasks requires participants to distinguish between the perspective experienced in the recollection and the event in the past of which it provides a perspective (e.g., Perner, Kloo & Stöttinger, 2007; Perner, Kloo & Rohwer, 2010). Of further interest, the overlap in brain activation for tasks presenting a perspective difference was clearly restricted to the left hemisphere. Activation in the right temporo-parietal junction was only found for the domain of theory of mind. This is compatible with the idea that the right temporo-parietal junction is specialized for processing of mental states (e.g., Saxe & Kanwisher, 2003; Saxe & Wechsler, 2005).

Meta-Analytic Results

A. Theory of mind with perspective (red), theory of mind without perspective (blue), and areas of overlap (purple).



B. Statistical comparisons of activations in A. Red: with > without perspective, blue: without > with perspective.



C. Theory of mind with perspective (red), episodic memory (blue), and areas of overlap (purple).

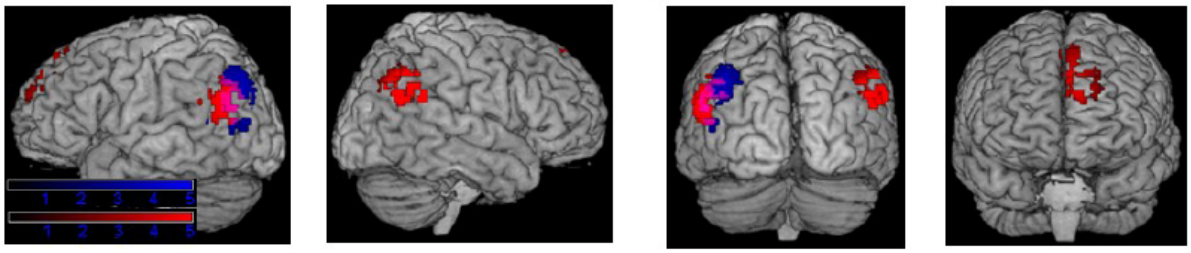


Figure 1. (A) Overlay of activation for theory of mind studies that did (red) or did not (blue) require processing of a perspective difference. (B) Results of a statistical comparison between activation for theory of mind tasks that did or did not require processing of a perspective difference. (C) Overlay of activation for theory of mind tasks that did require processing of a perspective difference and activation for episodic memory tasks. All maps are displayed at an uncorrected threshold of $p < .001$ and a minimum cluster extent of 10 voxels.

References

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage, 15*, 1059-68.

Mar, R. (2011). The neural basis of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103-34.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: the roles of left and right temporo-parietal junction. *Social Neuroscience, 1*, 245-258.

Perner, J., Kloo, D., & Gornik, E. (2007). Episodic memory development: Theory of mind is part of re-experiencing experienced events. *Infant & Child Development*, *16*, 471-490.

Perner, J., Kloo, D., & Rohwer, M. (2010). Retro- and Prospection for Mental Time Travel: Emergence of episodic remembering and mental rotation in 5- to 8-year old children. *Consciousness & Cognition, 19,* 802-815.

Perner, J. & Leekam, S. R. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology, 61*, 76-89.

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, 17*, 1451-72.

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

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 42*, 1435-1446.

Saxe, R., Young, L. (*in press*). Theory of Mind: How brains think about thoughts. In K. Ochsner & S. Kosslyn (eds.), *The Handbook of Cognitive Neuroscience*. Oxford University Press.

Radua, J., Mataix-Cols, D., Phillips, M.L., El-Hage, W., Kronhaus, D.M., Cardoner, N., & Surguladze, S. (2011). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry. in press*. http://dx.doi.org/10.1016/j.eurpsy.2011.04.001.

Counterfactual Reasoning and Reasoning with Beliefs:

Similar steps or separate paths?

Eva Rafetseder, University of Salzburg

Riggs, Peterson, Robinson, and Mitchell (1998) claimed that children have to apply counterfactual reasoning (CFR) when they answer false-belief questions. They used stories such as the Maxi-story: Maxi puts chocolate into the table drawer (location 1). He then goes to the playground. In his absence, his mother uses a piece of chocolate for her cake but then does not put it back into the drawer but puts it into the cupboard (location 2). Children are then asked a false-belief question ("Where will Maxi look for his chocolate?") and a counterfactual question ("If mother had not baked a cake, where would the chocolate be?"). Performance was highly correlated and children solved significantly more counterfactual questions than false belief questions which is compatible with the suggestion that counterfactual reasoning is a prerequisite for understanding false belief based actions.

Recent findings, however, indicate that children's answers to counterfactual questions may not be based on CFR but they give the correct answers using basic conditional reasoning (BCR: Rafetseder, Cristi-Vargas, & Perner, 2010), i.e., applying universally quantified conditionals that express general regularities. Use of CFR tends to emerge not before the age of 6 years.

Two studies investigated how children (7-14 years) perform on false-belief tasks when controlling for answers based on BCR. Both studies found a highly significant correlation (r = .58) between children using CFR to answer the counterfactual question and their correct answers to the false belief question. Moreover, hardly any children gave correct answers to the belief and wrong answers to the counterfactual question, suggesting that counterfactual reasoning is a prerequisite for predicting actions based on false beliefs.

This tight link between children's ability to answer counterfactual questions and identifying an agent's false belief suggests strongly that the "simulation gap" in theory use is filled by simulation but rather by counterfactual or, in the case of younger children, basic conditional reasoning. Perner & Roessler (2010) proposed that our basic understanding of action is based on objective rationality, i.e., teleology (similar to Csibra & Gergely 1998): agents take that action that brings about the goal (that which is desirable). This basic teleology is however limited. It cannot rationalise action when false belief or subjective desires are involved. To capture these cases, Perner and Roessler suggest, teleology is used counterfactually within the agent's perspective ("teleology in perspective"): "If the chocolate were in its old place and the goal is that Max should get the chocolate then he should go to its original place". By reasoning counterfactually about what should be done we use our own world knowledge without relying on simulation. We use our real suppositional reasoning without need to resort to any pretence reasoning or beliefs, and we need not introspect on beliefs or action tendencies. Rather, whatever is the case in the counterfactual scenario is what the agent believes (Gordon, 1995) and whatever needs doing in that scenario is what the agent is likely to do (Perner & Roessler 2010).

References

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*(2), 255–259.

Gordon, R. M. (1995). Simulation without introspection or inference from me to you. In M. Davies, & T. Stone (Eds.), *Mental simulation: Evaluations and applications* (pp. 53–67). Oxford: Blackwell.

Perner, J. & Roessler, J. (2010). Teleology and Causal Reasoning in Children's Theory of Mind. In J. Aguilar & A.A. Buckareff (Eds.), *Causing Human Action: New*

Perspectives on the Causal Theory of Action (chapter 14, 199–228). Cambridge, MA: Bradford Book, The MIT Press.

Riggs, K. J., Peterson, D. M., Robinson, E. J. & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13*(1), 73–90.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development, 81*(1), 376–389.

Representing Beliefs, Perspectives and Counterfactuals: A Puzzle

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

February 27, 2012

Abstract

What is involved in representing belief? This talk starts with three claims that are separately defensible but appear to be jointly inconsistent:

1. Infants can represent false beliefs from around their first birthday or earlier.
2. Being able to represent false beliefs involves being able to (i) process perspective differences or (ii) reason counterfactually (or both).
3. Infants cannot (i) process perspective differences nor (ii) engage in counterfactual reasoning until they are at least one year old.

The first part of the talk draws on research presented elsewhere in this symposium to defend each claim in turn. If each of the three claims (1)-(3) is true, they cannot be inconsistent after all. The second part of the talk explores ways of avoiding inconsistency by distinguishing measurement schemes and representational systems. There may be multiple ways of representing belief, and different ways of representing belief may vary in how flexible they are and what their implementation requires.

1. Introduction

The broader challenge is to understand what mindreading is, how it develops and which cognitive processes underpin it. One approach is to identify components of mindreading. As I understand it, Eva's and Matthias' research does just this. Eva argues that mindreading, at least insofar as it involves representing beliefs, may depend on counterfactual reasoning too.

And Matthias supports the claim that mindreading involves a separable ability to process perspective differences by showing that the neural correlates of this ability are found in both mindreading and non-mindreading tasks.

Eva's and Matthias' interpretations of their findings each appear to be in conflict with a widely held view about infant mindreading, as Eva notes in the paper on which her talk is based. These three claims appear to be jointly inconsistent:

1. Infants can represent false beliefs from around their first birthday or earlier.

2. Being able to represent false beliefs involves being able to (i) process perspective differences or (ii) reason counterfactually (or both).[1]

3. Infants cannot (i) process perspective differences nor (ii) engage in counterfactual reasoning until they are at least one year old.

Since these claims appear to be jointly inconsistent, it seems we must reject at least one. But which?

2. The puzzle

Can we reject (1), the claim that infants can represent false beliefs? This is what Eva suggests in her paper. But there is a growing body of evidence which appears to support (1). From around their first birthday infants predict actions of agents with false beliefs about the locations of objects (Onishi & Baillargeon 2005; Southgate et al. 2007) and choose different ways of interacting with others depending on whether their beliefs are true or false (Buttelmann et al. 2009; Knudsen & Liszkowski 2011; 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). The variety of paradigms measures—looking time, anticipatory looking, pointing and helping—makes it hard to dismiss these findings on methodological grounds. And while some have proposed that infants might be tracking behaviour only Perner & Ruffman (2005); Ruffman & Perner (2005), this proposal faces several objections (e.g. Song et al. 2008). The core issue is not whether hypothetical behavioural strategies might in principle explain theory of mind abilities; it is whether all of these subjects' actual behaviour

[1] Here and below I'm save words by not carefully distinguishing developmental claims about what acquiring an ability to represent false belief requires (Eva's main concern) from cognitive claims about what exercising an ability to represent false belief requires (Matthias' main concern).

reading capacities are flexible enough to explain the full range of their theory of mind abilities (Apperly & Butterfill 2009) In my view this is unlikely. So straightforward rejection of (1) is not an option.

Can we reject (2)? Since Matthias' and Eva's talks are (in part) arguments for this claim, let me come back to it after first considering whether we can reject claim (3).

Can we reject (3), the claim that infants neither process perspective differences nor engage in counterfactual reasoning until they are at least one year old? One difficulty is that current developmental evidence almost uniformly supports this claim (Rafetseder et al. 2010; Beck & Guthrie 2011). A second problem is that the development of abilities to reason counterfactually,[2] like the development of abilities to represent false belief, appears to involve working memory and inhibitory control (on counterfactuals: Drayton et al. 2011; Beck et al. 2011; on false beliefs: Apperly et al. 2008, 2009; Lin et al. 2010; McKinnon & Moscovitch 2007 experiments 4-5; Saxe et al. 2006). As even committed nativists are likely to agree, capacities for inhibitory control and working memory develop over several years and are limited in infants (e.g. Carlson 2005). So we cannot straightforwardly reject (3).

Since we can't straightforwardly reject (1) or (3), let's consider whether (2) can be rejected. This is the claim that being able to represent false beliefs involves being able to process perspective differences or to reason counterfactually. One might attempt to reject this claim on the grounds that the link between false belief, perspective differences and counterfactual reasoning is spurious and a consequence of *extraneous demands* accompanying standard false belief tasks. Here the idea is that, although certain false belief *tasks* impose demand processing perspective differences or core components of counterfactual reasoning, these demands are extraneous in the sense that they are not a consequence merely of being required to represent false beliefs (proponents of this view include Carpenter et al. 2002, p. 417, Bloom & Gelman 2000, and Leslie & Polizzi 1998). I want to go slowly here and consider two versions of this idea.

Sometimes it is taken for granted that representing false beliefs does not necessarily and intrinsically involve inhibition; and so the fact that success on many false belief tasks depends on inhibitory control is interpreted as revealing a defect of the task. Someone who holds this view will interpret the evidence Eva and Matthias present as showing that the false belief tasks they consider are defective measures of the ability to represent false beliefs. But I think it is a mistake to assume that representing false belief—or, indeed, representing any propositional attitude—does not intrinsically demand cog-

[2] Here and throughout I follow Eva Rafetseder in using the term 'counterfactual reasoning' to refer to reasoning which essentially involves comparing two possible situations; the term is sometimes used more broadly to include the sorts of representations involved in pretence.

nitive resources such as inhibitory control. On any standard view, propositional attitudes such as beliefs form complex causal structures, have arbitrarily nest-able 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 there is anything representing which should consume scarce cognitive resources it is surely states with this combination of properties. So there are sound theoretical reasons to suppose that representing false beliefs could intrinsically require inhibitory control and working memory (see also Russell 1999). And of course Eva and Matthias can provide theoretical reasons for supposing that representing false beliefs involves processing perspective differences and a core component of genuine counterfactual reasoning (Perner et al. 2007). This is why we should not take for granted that representing belief can be separated from processing perspective differences and counterfactual reasoning; there is no good theoretical reason to assume that the connections must reflect extraneous task demands.

A less obviously flawed approach to rejecting (2) involves analysing particular false belief tasks. Two explore this idea it will be useful to distinguish between two types of false belief task, call them Category A and Category B. By stipulation Category A tasks have these features:

- Children tend to pass them some time after their third birthday.
- Abilities to pass these tasks has a protracted developmental course stretching over months if not years.
- Success on these tasks is correlated with developments in executive function (Perner & Lang 1999; Sabbagh 2006) and language (Astington & Baird 2005).
- Success on these tasks is facilitated by explicit training (Slaughter & Gopnik 1996) and environmental factors such as siblings (Clements et al. 2000a; Hughes & Leekam 2004).
- Abilities to succeed on these tasks typically emerge from extensive participation in social interactions (as Hughes et al. 2006 suggest).

All the other false belief tasks are in Category B. It is these Category B tasks on which infants succeed. The proposal we are considering (and which I shall reject) is this.

> All Category A tasks impose a requirement (or set of requirements) other than the requirement to represent a false belief; call this the *Extraneous Requirement*. The connections between success on Category A tasks and processing perspective differences and counterfactual reasoning are a consequence just of the extraneous requirement.

I am going to argue that this proposal should be rejected. First note that the proposal is harder to defend than sometimes assumed. For instance, philosophers sometimes mistakenly assume that all Category A tasks involve language or communication, whereas Category B tasks do not. But some Category B tasks involve language and communication (Knudsen & Liszkowski 2011; Song et al. 2008) and there are non-verbal tasks in Category A (Call & Tomasello 1999; Low 2010 Study 2). We should also note that children who fail Category A tasks can answer questions about perception or pretence that are word-for-word identical with the questions about false belief that they cannot answer correctly (Gopnik et al. 1994; see also?). So the Extraneous Requirement cannot be straightforwardly linked to language or communication. There is a more general problem facing attempts to identify the Extraneous Requirement. There is much variation among the tasks in Category A. It is not just that some are verbal whereas others are nonverbal. It is also that some involve prediction whereas others involve retrodiction or justification (e.g. Wimmer & Mayringer 1998), some concern the first-person perspective, whereas others involve a second- or third-person perspective (e.g. Gopnik & Slaughter 1991), some involve interaction whereas in others the subject is a mere observer (e.g. Chandler et al. 1989), and some involve prediction actions whereas others involve predicting desires (Astington & Gopnik 1991) or selecting an argument appropriate for someone with a false belief (Bartsch & London 2000). Despite all this variation and more, the Category A tasks all appear to measure a developmental transition (Wellman et al. 2001). This is a significant obstacle to identifying the Extraneous Requirement. In fact, my view the currently available evidence supports the conclusion that there is no Extraneous Requirement. And as far as I know the only plausible candidate for the transition Category A tasks measure is a transition in the ability to represent beliefs. This is why I don't think that we can straightforwardly reject (2).

To sum up so far, I have suggested that there are good reasons for each of (1)-(3) and argued that none of these claims can be straightforwardly rejected. This leads to a puzzle. Because (1)-(3) appear to be jointly inconsistent, it seems that one of them must be rejected. This is puzzling because (1)-(3) also appear to be individually defensible. None of the arguments I have offered are decisive; what have aimed to establish is just that the question of which claim to reject is to a significant degree puzzling. And although various researchers have taken one side or another, I think it's fair to say that no one has yet succeeded in resolving the puzzle.

3. Distinguishing implicit and explicit isn't enough

Could we resolve the puzzle by distinguishing between implicit and explicit representations of false belief? Given such a distinction we could remove the contradiction by making the following modifications:

> **1′**. Infants can represent false beliefs *implicitly* from around their first birthday or earlier.
>
> **2′**. Being able to represent false beliefs *explicitly* involves being able to (i) process perspective differences or (ii) reason counterfactually (or both).
>
> **3′**. Infants cannot (i) process perspective differences nor (ii) engage in counterfactual reasoning until they are at least one year old.

When the primary measure of early false belief understanding was looking times or anticipatory eye movements, this seemed promising (Clements & Perner 1994; Clements et al. 2000b; Garnham & Ruffman 2001; Ruffman et al. 2001). However, as already mentioned, we now know that one-year-old infants succeed on false belief tasks which involve actively helping others, interpreting their utterances and pointing to provide information (Buttelmann et al. 2009; Knudsen & Liszkowski 2011; Southgate et al. 2010), and that older children still fail some nonverbal false belief tasks (Call & Tomasello 1999). All other things being equal, it would be natural to suppose that success on these sorts of task reflects explicit representations of false belief. Certainly we lack a theoretically coherent and independently empirically motivated account of the distinction between implicit and explicit representation that associates exactly those tasks infants succeed on (strictly: Category B tasks) with implicit representations of false belief. So we don't know that distinguishing implicit from explicit representations does enable us to avoid contradiction.

Of course I am not denying that, considered as a purely formal move, distinguishing between implicit and explicit representations could in principle enable us to avoid contradiction. The problem is to understand what this distinction amounts to clearly enough to explain the existing data and generate novel predictions (Apperly 2010, p. 154).

4. How to resolve the puzzle: first step

I can't fully resolve the puzzle but I do think we can make some progress towards resolving it. This section describes the first of two steps towards resolution.

Let me start with an observation. When philosophers discuss what belief is, they generally disagree. They disagree about what sorts of contents beliefs have (for instance about whether their contents are propositional or map-like), about which norms and causal roles characterise the attitude of believing (for instance about whether part of being a belief is being a state that ought to be true) and about what it is for a belief to be true or false (for instance about whether truth or falsity is merely a matter of the belief's value as a guide to action). But when researchers focus on representing belief, it seems that no notice is taken of these disagreements. I think this is a mistake. Different views on what belief is are related to different accounts of what it is to represent belief. After all, what is involved in accurately representing a false belief depends in part on what a belief is and what it is for a belief to be false.

A second observation is that there is no a priori reason to assume that all subjects who represent beliefs represent them with perfect accuracy. Suppose, just for the sake of argument, that beliefs are relations to full-blown Fregean propositions. A particular subject might nevertheless represent beliefs as relations to mere singular propositions. This subject's abilities to track others' beliefs, and to predict their actions, would be limited in various ways (Salmon 1986). But despite these limits, having such an ability to represent beliefs may still be useful in a wide range of everyday circumstances. In fact, in many cases this subject might be indistinguishable from another subject with a more accurate conception of belief. An analogy may help (this is based on Matthews 2007). There are various ways to represent the efficiency of your car. If you were very exact, you might use a measure relating force to fuel consumption. But many people will instead measure distance in relation to fuel consumption, or even just fuel consumption in relation to periods of time (perhaps monthly). The point is that we have different ways of representing (measuring) efficiency. Some of these are highly flexible in the sense that they can cope with variations in landscape and even gravity but they are also more costly to implement in terms of the time and background knowledge they demand; while others are less costly to implement but also more limited in the sense that they only work in a more limited range of circumstances (perhaps only when the car is used in roughly the same way each month). What I am suggesting is that there is a parallel for representing (measuring) belief. On rich conceptions of belief, accurately representing belief will enable one to make fine-grained distinctions and to track complex contents, and this will demand conceptual sophistication and impose cognitive demands. On other, simpler conceptions of belief, accurately representing belief will yield less flexibility in distinguishing what others belief and in predicting their actions but the reduced flexibility in greater cognitive efficiency.

Whatever belief actually is, there may be multiple ways of representing

it, just as there are multiple ways of measuring fuel efficiency. Different ways of representing belief will vary in how flexible they are and how demanding they are to implement.

Saying that a subject can represent beliefs is like saying that she can represent fuel efficiency. It leaves open the question of how she represents beliefs. For instance, does she represent the contents of beliefs as simple tuples, as propositions of some kind, as images or maps, or what? The answer may vary for different subjects. And of course a single subject may use multiple ways of representing beliefs in different contexts (much as a single subject might rely on different ways of measuring fuel efficiency in different contexts). It is striking that almost no research has directly addressed the issue of how beliefs are represented in infants, children or adults.

At this point it is perhaps tempting to suppose that we already have everything we need to resolve the puzzle. The idea would be this:

> **1″**. Infants can represent false beliefs *using a comparatively simple measure* from around their first birthday or earlier.
>
> **2″**. Being able to represent false beliefs *using a comparatively sophisticated measure* involves being able to (i) process perspective differences or (ii) reason counterfactually (or both).
>
> **3″**. Infants cannot (i) process perspective differences nor (ii) engage in counterfactual reasoning until they are at least one year old.

I think all these claims are true, and with Ian Apperly I have developed in detail a conjecture about the comparatively simple measure infants and others may use in representing beliefs (Butterfill & Apperly 2011). It turns out that different measures of belief generate testable and novel predictions about the limits of subjects' abilities to track beliefs, so although we haven't tested the conjecture yet it is a testable conjecture.

But although this is my view and I stand by it, I don't think it resolves the puzzle all by itself. The problem is this. Success on many standard false belief tasks in Category A is in principle possible using only a very simple measure for representing beliefs. Further, Category A tasks do not seem to differ from Category B tasks with respect to which measures for representing beliefs could in principle explain success on these tasks. So the view raises and does not answer the question of why children continue to systematically fail false belief tasks in Category A years after they can pass false belief tasks in Category B.

5. How to resolve the puzzle: second step

Here is another observation. In many domains it is widely recognised that patterns of performance in human subjects are best explained on the hypothesis that multiple representational systems are involved. This is true for representations of faces (Morton & Johnson 1991), of phonemes and of causes (Butterfill 2009), for instance. It is also true for representations of number (e.g. Feigenson & Carey 2003), agency and space (Carey & Spelke 1996). In these domains and others, it seems we have to distinguish representations that paradigmatically feature in reflective thought and talk from representations that feature in modules or core systems. These modules are typically functional early in development and throughout most of the life course, make few or no demands on scarce cognitive resources like working memory, and are limited and inflexible with respect to the information they can process and output.

There are some striking parallels between patterns of performance which seem to license distinguishing multiple systems for representing number and patterns of performance in the case of reasoning about false beliefs and mental states more generally (Apperly & Butterfill 2009). This suggests that, in humans at least, representations of belief may feature in multiple cognitive systems.

When a subject performing a task is described as representing a number or a belief, this leaves open which system is involved. Is it a relatively efficient system or one that places demands on working memory or inhibitory control? When does the system appear in development? In short, descriptions of subjects as representing beliefs need to be relativised in a second way. They need to be relativised not just to a measurement scheme (see section 4 on page 6) but also to a system.

The two kinds of relativisation—to a measurement scheme and to a system—are plausibly related. A representational system can only be cognitively efficient by avoiding sophisticated measurement schemes, which will result in limits on flexibility; equally, a representational system can only achieve flexibility in representing beliefs by implementing a sophisticated measurement scheme, and this will plausibly impose cognitive overheads in many situations.

So here is a final refinement of the claims which make up the puzzle that I started with:

> **1‴**. Infants can represent false beliefs *using a comparatively simple measure* and *in a modular process* from around their first birthday or earlier.
>
> **2‴**. Being able to represent false beliefs *using a comparatively sophisticated measure* and *in a non-modular process* involves

being able to (i) process perspective differences or (ii) reason counterfactually (or both).

3‴. Infants cannot (i) process perspective differences nor (ii) engage in counterfactual reasoning until they are at least one year old.

At this point the claims are no longer contradictory. But it remains an open question whether we now have all the ingredients necessary to explain both early success on Category B false belief tasks and late failure on Category A tasks or whether some further complexity in the notion of representing belief also needs to be added to the mix.[3]

6. Conclusion

In conclusion, I have argued that Eva's and Matthias' findings generate a puzzle resolving which may require confronting the question, What is a representation of belief?

References

Apperly, I. A. (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Hove: Psychology Press.

Apperly, I. A., 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*, 1093–1108.

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.

Astington, J. & Gopnik, A. (1991). Developing understanding of desire and intention. In A. Whiten (Ed.), *Natural Theories of the Mind: evolution, development and simulation of everyday mindreading* (pp. 39–50). Oxford: Blackwell.

[3] Apperly (2010, p. 155-6) appears to suggest a solution additionally involving egocentric bias, but I'm not yet fully convinced.

Bartsch, K. & London, K. (2000). Children's use of mental state information in selecting persuasive arguments. *Developmental Psychology*, *36*(3), 352–365.

Beck, S. R., Carroll, D. J., Brunsdon, V. E., & Gryg, C. K. (2011). Supporting children's counterfactual thinking with alternative modes of responding. *Journal of Experimental Child Psychology*, *108*(1), 190–202.

Beck, S. R. & Guthrie, C. (2011). Almost thinking counterfactually: Children's understanding of close counterfactuals. *Child Development*, *82*(4), 1189–1198.

Bloom, P. & Gelman, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*, B25–B31.

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

Butterfill, S. (2009). Seeing causes and hearing gestures. *Philosophical Quarterly*, *59*(236), 405–428.

Butterfill, S. & Apperly, I. A. (submitted 2011). How to construct a minimal theory of mind. `http://butterfill.com/papers/minimal_theory_of_mind/`.

Call, J. & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, *70*(2), 381–395.

Carey, S. & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515–533.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616.

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

Chandler, M., Fritz, A., & Hala, S. (1989). Small scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development*, *60*, 1263–1277.

Clements, W. & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395.

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

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

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

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

Drayton, S., Turley-Ames, K. J., & Guajardo, N. R. (2011). Counterfactual thinking and false belief: The role of executive function. *Journal of Experimental Child Psychology*, *108*(3), 532–548.

Feigenson, L. & Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584.

Garnham, W. & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding or belief. *Developmental Science*, *4*(1), 94–100.

Gopnik, A. & Slaughter, V. (1991). Young children's understanding of changes in their mental states. *Child Development*, *62*, 98–110.

Gopnik, A., Slaughter, V., & Meltzoff, A. (1994). Changing your views: How understanding visual perception can lead to a new theory of mind. In C. Lewis & P. Mitchell (Eds.), *Children's Early Understanding of Mind: origins and development*. Hove: Erlbaum.

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.

Knudsen, B. & Liszkowski, U. (forthcoming 2011). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*.

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.

Leslie, A. & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*(2), 247–253.

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.

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

Matthews, R. (2007). *The measure of mind: propositional attitudes and their attribution*. Oxford: Oxford University Press.

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.

Morton, J. & Johnson, M. H. (1991). Conspect and conlern: A two-process theory of infant face recognition. *Psychological Review*, *98*(2), 164–181.

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

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

Perner, J., Rendl, B., & Garnham, A. (2007). Objects of desire, thought, and reality: Problems of anchoring discourse referents in development. *Mind & Language*, *22*(5), 475–513.

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

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "Nearest possible world". *Child Development*, *81*(1), 376–389.

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*, 201–224.

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

Russell, J. (1999). Cognitive development as an executive process - in part: a homeopathic dose of piaget. *Developmental Science*, *2*(3), 247–295.

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

Salmon, N. (1986). *Frege's Puzzle*. Cambridge, Mass: MIT Press.

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*(3-4), 284–298. PMID: 18633794.

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

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.

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.

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.