

# False Beliefs, Perspectives and Counterfactuals

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

February 26, 2012

## Abstract

\*\*\*

## 1. Introduction

The 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.

There appears to be a conflict between Eva's and Matthias' interpretations of their findings and a widely held view about infant mindreading (as Eva notes in the paper on which her talk is based). Here are three claims which are collectively 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.

Since these claims are collectively inconsistent, 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 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 abilities to reason counterfactually,<sup>1</sup> like abilities to represent false belief, appear 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). 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

---

<sup>1</sup> 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.

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 cognitive 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. To 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. 2000; 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 (Wimmer & Mayringer 1998, e.g.), some concern the first-person perspective, whereas others involve a second- or third-person perspective (Gopnik & Slaughter 1991, e.g.), some involve interaction whereas in others the subject is a mere observer (Chandler et al. 1989, e.g.), 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) are collectively inconsistent, one of them must be rejected. The puzzle is: Which should we reject? 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. How to resolve the puzzle

## References

- 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.
- 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.
- Call, J. & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, 70(2), 381–395.
- 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., Rustin, C., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. *Developmental Science*, 3(1), 81–92.
- 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.
- 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). Co-operation 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.
- 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.
- 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. & 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.
- 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.