

Origins of Mind

Lecture 07

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

November 22, 2016

title-slide

Knowledge of Mind

slide-3 The challenge is to explain the developmental emergence of mindreading.

Let me explain.

slide-4 *Mindreading* is the process of identifying mental states and purposive actions as the mental states and purposive actions of a particular subject.

slide-5 Researchers sometimes use the term ‘theory of mind’.

‘In saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and to others’ (Premack and Woodruff 1978, p. 515)

slide-6 So, to be clear about the terminology, to have a theory of mind is just to be able to mindread, that is, to identify mental states and purposive actions as the mental states and purposive actions of a particular subject.

slide-7 So the challenge is to explain the emergence of mindreading. You know (let’s say) that Ayesha believes Beatrice is in the library. Humans are not born knowing individuating facts about others’ beliefs. How do they come to be in a position to know such facts? Meeting this challenge initially seems simple. But, as you’ll see, we quickly end up with a puzzle. I think this puzzle requires us to rethink what is involved in having a conception of the mental.

slide-8 I shall focus on awareness of others’ beliefs to the exclusion of other mental states.

There's no theoretical reason for this; it's just a practical thing.

And what we learn about belief will generalise to other mental states.

slide-9 How can we test whether someone is able to ascribe beliefs to others? Here is one quite famous way to test this, perhaps some of you are even aware of it already. Let's suppose I am the experimenter and you are the subjects. First I tell you a story ...

In a standard *false belief task*, '[t]he subject is aware that he/she and another person [Maxi] witness a certain state of affairs x. Then, in the absence of the other person the subject witnesses an unexpected change in the state of affairs from x to y' (Wimmer and Perner 1983, p. 106). The task is designed to measure the subject's sensitivity to the probability that Maxi will falsely believe x to obtain.

slide-13 Here's the really surprising thing.

Children do really badly on this until they are around four years of age.

And they seem to develop the ability to pass this task only gradually, over months or years.

(There's something else that isn't surprising to most people but should be: adult humans not only nearly always provide the answer we're calling 'correct': they also believe that there is an obviously correct answer and that it would be a mistake to give any other answer. I'll return to this point later.)

(NB: The figure is not Wimmer & Perner's but drawn from their data.)

slide-14 There's been some stuff in the press recently about bad science, mainly some dodgy methods and failures to replicate.

slide-15 So you'll be pleased to know that a meta-study of 178 papers confirmed Wimmer & Perner's findings.

Now there is clearly some variation here.

That's because different researchers implemented different versions of the original task.

We can use the meta-analysis of these experiments as a shortcut to finding out what sorts of factors affect children's performance.

slide-16 One factor that seems to make hardly any difference is whether you ask children about others' beliefs or their own beliefs.

To repeat, you get essentially the same results whether you ask children about others' beliefs or their own beliefs.

Children literally do not know their own minds.

slide-17 What happens if we involve the child by having her interact with the protagonist?

The task becomes easier for children of all ages, but the transition is essentially the same (participation does not interact with age Wellman et al. 2001, pp. 665-7).

slide-18 Finally, although there are some cultural differences, you get the same transition in seven different countries.

slide-19 So our challenge was to explain the emergence of mindreading. At this point, up until around, it seemed quite straightforward to most researchers. We seemed to know that children are unaware of mental states until around four years. And a lot of studies looked at which factors affect their acquiring this awareness. These studies showed that executive function, language and rich forms of social interaction are all important. All of this supported something like the story that Sellars tells in his famous Myth of Jones.

slide-20 *todo*: describe Sellars' myth; link to Gopnik theory theory idea.

slide-21 But there was a big surprise in store for us.

unit_411

Infants Track False Beliefs

slide-23 How can we test whether someone is able to ascribe beliefs to others? Here is one quite famous way to test this, perhaps some of you are even aware of it already. Let's suppose I am the experimenter and you are the subjects. First I tell you a story ...

In a standard *false belief task*, '[t]he subject is aware that he/she and another person [Maxi] witness a certain state of affairs x. Then, in the absence of the other person the subject witnesses an unexpected change in the state of affairs from x to y' (Wimmer and Perner 1983, p. 106). The task is designed to measure the subject's sensitivity to the probability that Maxi will falsely believe x to obtain.

Recall the experiment that got us started.

These experimenters added an anticipation prompt and measured to which box subjects looked first (Clements and Perner 1994).

(Actually they didn't use this story; theirs was about a mouse called Sam and some cheese, but the differences needn't concern us.)

slide-26 What got me hooked philosophical psychology, and on philosophical issues in the development of mindreading in particular was a brilliant finding by Wendy Clements who was Josef Perner's phd student.

These findings were carefully confirmed (Clements et al. 2000; Garnham and Ruffman 2001; Ruffman et al. 2001).

Around 2000 there were a variety of findings pointing in the direction of a conflict between different measures.

These included studies on word learning (Carpenter et al. 2002; Happé and Loth 2002) and false denials (Polak and Harris 1999).

But relatively few people were interested until ...

unit_416

Mindreading: a Developmental Puzzle

slide-56 The challenge is to explain the emergence, in evolution or development, of mindreading. Initially it looked like this was going to be relatively straightforward and involve just language, social interaction and executive function. So a Myth of Jones style story seemed viable. But the findings of competence in infants of around one year of age changes this. These findings tell us that not all abilities to represent others' mental states can depend on things like language. And, as I've been stressing, these findings also create a puzzle. The puzzle is, roughly, how to reconcile infants' competence with three-year-olds' failure.

***todo*:** There are at least two possible puzzles you might focus on: 1. How can we avoid the apparent contradiction in the evidence? 2. How can we explain the apparent discrepancy between infants and 3-year-olds' performances? Minimal theory of mind might resolve puzzle 1 but it won't by itself resolve puzzle 2. This wasn't clear enough in the lecture (sorry!).

slide-60 Let me start by taking you back to the early eighties. (Has anyone else been enjoying Deutschland drei und achtzig?)

3-year-olds fail false belief tasks

slide-61 3-year-olds fail a wide variety of tasks where they are asked about a false belief, or asked to predict how someone with a false belief will act, ...

slide-62 or asked to predict what someone with a false belief will desire ...

slide-63 ... or to retrodict or explain a false belief after being shown how someone acts.

slide-64 Further, lots of factors make no difference to 3-year-olds' performance: they fail tasks about other's beliefs and they fail tasks about their own beliefs;

slide-65 they fail when they are merely observers as well as when they are actively involved;

slide-66 they fail when a verbal response is required and also when a non-verbal communicative response or even a noncommunicative response (such as hiding an object) is required.

slide-67 And they fail test questions which are word-for-word identical to desire and pretence tasks

slide-68 An A-Task is any false belief task that children tend to fail until around three to five years of age.

slide-69 Why do children systematically fail A-tasks? There is a simple explanation ...

Children fail

because they rely on a model of minds and actions that does not incorporate beliefs

[Stress that, on this view, children do have a model of minds and actions. It's just that it doesn't incorporate belief.] Perner and others have championed the view that children who failed A-tasks lack a metarepresentational understanding of propositional attitudes altogether. But this view has recently (well, not that recently, it's nearly a decade old now) been challenged by Hannes' discovery that children can solve tasks which are like A-tasks but involve incompatible desires (Rakoczy et al. 2007). I think this makes plausible the thought that there is an age at which children fail A-tasks not because they have a problem with mental states in general, but because they have a problem with beliefs in particular.

slide-70 It turns out that, in the first and second years of life, infants show abilities to track false beliefs on a variety of measures.

slide-75 A non-A-Task is a task that is not an A-task. I was tempted to call these B-Tasks, but that would imply that they have a unity. And whereas we know from a meta-analysis that A-tasks do seem to measure a single

underlying competence, we don't yet know whether all non-A-tasks measure a single thing or whether there might be several different things.

slide-76 Why do infants systematically pass non-A-tasks in the first year or two of life? There is a simple explanation ...

And this is *almost* everything we need to generate the puzzle about development.

slide-77 Children fail A-tasks because they rely on a model of minds and actions that does not incorporate beliefs.

Children pass non-A-tasks by relying on a model of minds and actions that does incorporate beliefs.

The dogma of mindreading: any individual has at most one model of minds and actions at any one point in time.

We've seen that ... To get a contradiction we need one further ingredient.

slide-78 These three claims are jointly inconsistent so one of them must be false. Researchers disagree about which claim to reject. But I suppose you can tell from how I've labelled them which one I propose to reject.

slide-80 The puzzle is a little bit like the puzzle we had in the case of knowledge of physical objects. But it's also different. In the case of physical objects, the conflict was between measures involving looking and measures involving searching. In this case it's different, because on the infant side there is not just looking but also acting (e.g. helping) and even communicating.

slide-87 Children fail A-tasks because they rely on a model of minds and actions that does not incorporate beliefs.

Children pass non-A-tasks by relying on a model of minds and actions that does incorporate beliefs.

The dogma of mindreading: any individual has at most one model of minds and actions at any one point in time.

How does this bear on the case of mindreading?

slide-88 Conjecture

Infants have core knowledge of minds and actions.

Core knowledge is sufficient for success on non-A-tasks.

Infants lack knowledge of minds and actions.

Knowledge is necessary for success on A-tasks.

Why accept this conjecture? So far no reason has been given at all. And it barely makes sense. There are just so many assumptions. All I'm really saying is that I hope this case, knowledge of minds, will turn out to be like the other cases.

slide-89 Why accept this conjecture? And what form does the core knowledge take. In every case so far, we have had to identify infant with adult competencies. (Core knowledge is for life, not just for infancy.)

unit_421

Mindreading in Adults: Dual Processes

slide-91 A process involves *belief-tracking* if how processes of this type unfold typically and nonaccidentally depends on facts about beliefs. So belief tracking can, but need not, involve representing beliefs.

belief-tracking is sometimes but not always automatic

slide-92 A process is *automatic* to the degree that whether it occurs is independent of its relevance to the particulars of the subject's task, motives and aims.

slide-93 Or, more carefully, does belief tracking in human adults depend only on processes which are automatic?

There is now a variety of evidence that belief-tracking is sometimes and not always automatic in adults. Let me give you just one experiment here to illustrate.

slide-105 [skip this slide]

One way to show that mindreading is automatic is to give subjects a task which does not require tracking beliefs and then to compare their performance in two scenarios: a scenario where someone else has a false belief, and a scenario in which someone else has a true belief. If mindreading occurs automatically, performance should not vary between the two scenarios because others' beliefs are always irrelevant to the subjects' task and motivations.

slide-106 [skip this slide]

Schneider et al. (2012) did just this. They showed their participants a series of videos and instructed them to detect when a figure waved or, in a second

experiment, to discriminate between high and low tones as quickly as possible. Performing these tasks did not require tracking anyone's beliefs, and the participants did not report mindreading when asked afterwards.

on experiment 1: 'Participants never reported belief tracking when questioned in an open format after the experiment ("What do you think this experiment was about?"). Furthermore, this verbal debriefing about the experiment's purpose never triggered participants to indicate that they followed the actor's belief state' (Schneider et al. 2012, p. 2)

Nevertheless, participants' eye movements indicated that they were tracking the beliefs of a person who happened to be in the videos.

In a further study, Schneider et al. (2014) raised the stakes by giving participants a task that would be harder to perform if they were tracking another's beliefs. So now tracking another's beliefs is not only irrelevant to performing the tasks: it may actually hinder performance. Despite this, they found evidence in adults' looking times that they were tracking another's false beliefs. This indicates that 'subjects ... track the mental states of others even when they have instructions to complete a task that is incongruent with this operation' (Schneider et al. 2014, p. 46) and so provides evidence for automaticity.¹

Further evidence that mindreading can occur in adults even when counterproductive has been provided by Kovács et al. (2010), who showed that another's irrelevant beliefs about the location of an object can affect how quickly people can detect the object's presence, and by van der Wel et al. (2014), who showed that the same can influence the paths people take to reach an object. Taken together, this is compelling evidence that mindreading in adult humans sometimes involves automatic processes only.

slide-107 belief-tracking is sometimes but not always automatic

slide-112 Children fail A-tasks because they rely on a model of minds and actions that does not incorporate beliefs.

Children pass non-A-tasks by relying on a model of minds and actions that does incorporate beliefs.

¹ It is hard to completely rule out the possibility that belief tracking is merely spontaneous rather than automatic. I take the fact that belief tracking occurs despite plausibly making subjects' tasks harder to perform to indicate automaticity over spontaneity. If non-automatic belief tracking typically involves awareness of belief tracking, then the fact that subjects did not mention belief tracking when asked after the experiment about its purpose and what they were doing in it further supports the claim that belief tracking was automatic.

The dogma of mindreading: any individual has at most one model of minds and actions at any one point in time.

How might this help us with the puzzle about development?

slide-113 Recall this conjecture from earlier ...

Conjecture

Infants have core knowledge of minds and actions.

Core knowledge is sufficient for success on non-A-tasks.

Infants lack knowledge of minds and actions.

Knowledge is necessary for success on A-tasks.

The first challenge was to say what core knowledge of minds might be ...

core knowledge of minds = the representations underpinning automatic belief-tracking?

slide-114 The first challenge was to say what core knowledge of minds might be ...

slide-117 So far we have no evidence for the conjecture!

unit_441

Minimal Theory of Mind

unit_451

Signature Limits

slide-149 There is some evidence that this prediction is correct. Jason Low and his colleagues set out to test it. They have now published three different papers showing such limits; and Hannes Rakoczy and others have more work in progress on this. Collapsing several experiments using different approaches, the basic pattern of their findings is this ...

Take non-automatic responses first; in this case, communicative responses. When you do a false-belief-identity task, you see the pattern you also find for false-belief-locations tasks. But things look different when you measure non-automatic responses ...

The non-automatic responses all show the signature limit of minimal models of the mental. This is evidence for the hypothesis that Some automatic belief-tracking systems rely on minimal models of the mental.

I also hear that quite a few scientists have pilot data that speaks against this signature limit.

One particular task for future research will be to examine whether other automatic responses to scenarios involving false beliefs about identity, such as response times and movement trajectories, are also subject to this signature limit.

slide-150 Just say that you can do this with other stimuli and paradigms, and we have done this with infants and would like to do it with adults.

These findings complicate the picture: is helping driven by automatic processes only? If not, why do we predict that the signature limit of minimal theory of mind is found in this case too?

slide-152 Look at the three year olds. What might make us think that three year old's responses are a consequence of the same system that underpin's adults' automatic responses? One compelling consideration is that three year old's responses manifest to the same signature limit as adults'.

same signature limit -> same system

slide-153 Scott and colleagues (Scott et al. 2015) provided other evidence suggesting that infants' mindreading may be relatively sophisticated. Specifically, 17-month-olds watched a thief attempt to steal a preferred object (a rattling toy) when its owner was momentarily absent by substituting it with a less-preferred object (a non-rattling toy). Infants looked longer when the thief substituted the preferred object with a non-visually-matching silent toy compared to when the thief substituted it with a visually-matching silent toy. The authors postulated that infants can ascribe to the thief an intention to implant in the owner a false-belief about the identity of the substituted toy. The authors further suggested that infants make such ascriptions only when the substitution involves a visually-matching toy and the owner will not test whether the toy rattles on her return.

However, Scott et al.'s (Scott et al. 2015) explanations also require postulating that infants take the thief to be strikingly inept; despite having opportunity simply to pilfer from a closed box known to contain at least three rattling toys, the thief engages in elaborate deception which will be uncovered whenever the substituted toy is next shaken and the thief, as sole suspect, easily identified. A further difficulty is that factors unrelated to the thief's mental states vary between conditions, such as the frequencies with which toys visually matching one present during the final phase of the test trial have rattled. These considerations jointly indicate that further evidence would be needed to support the claim that humans' early mindreading capacity en-

ables them to ascribe intentions concerning false beliefs involving numerical identity.

slide-154 It has to be said that not everyone is convinced ..

‘the theoretical arguments offered [...] are [...] unconvincing, and [...] the data can be explained in other terms’ (Carruthers 2015b; see also Carruthers 2015a).

What is my response? Yes, the data can be explained in other terms, at least post hoc; and certainly there is as yet insufficient data for certainty. What about the theoretical arguments? Partners in crime defence ... theoretical arguments for multiple systems for belief are the same as the theoretical arguments for physical cognition or number cognition (but that’s a different talk).

slide-157 So let me conclude. The challenge we have been addressing was to understand the emergence of mindreading. Initially this seemed straightforward: you learn this from social interaction using language as a tool (compare Gopnik’s theory theory). However, the discovery that abilities to track beliefs exist in infants from around 7 months or earlier initially suggested a different picture: one on which mindreading was likely to involve core knowledge. But, as always, things are not so straightforward.

slide-158 Children fail A-tasks because they rely on a model of minds and actions that does not incorporate beliefs.

Children pass non-A-tasks by relying on a model of minds and actions that does incorporate beliefs.

The dogma of mindreading: any individual has at most one model of minds and actions at any one point in time.

The puzzle.

slide-159 Recall this conjecture from earlier ...

Conjecture

Infants have core knowledge of minds and actions.

Core knowledge is sufficient for success on non-A-tasks.

Infants lack knowledge of minds and actions.

Knowledge is necessary for success on A-tasks.

The first challenge was to say what core knowledge of minds might be ...

core knowledge of minds = the representations underpinning automatic belief-tracking?

slide-160 The first challenge was to say what core knowledge of minds might be ...

slide-163 So far we have no evidence for the conjecture!

slide-164 theme A: explain the origins of knowledge of others minds : development as rediscovery. There is a modular capacity (= core knowledge). But this doesn't lead to adult-like understanding for years, and the acquisition of adult-like understanding hinges on language; may involve completely different model of mental states.

References

- Carpenter, M., Call, J., and Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology*, 20:393–420.
- Carruthers, P. (2015a). Mindreading in adults: evaluating two-systems views. *Synthese*, forthcoming:1–16.
- Carruthers, P. (2015b). Two systems for mindreading? *Review of Philosophy and Psychology*, 7(1):141–162.
- Clements, W. and Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9:377–395.
- Clements, W., Rustin, C., and McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. *Developmental Science*, 3(1):81–92.
- Garnham, W. and Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding or belief. *Developmental Science*, 4(1):94–100.
- Happé, F. and Loth, E. (2002). 'theory of mind' and tracking speakers' intentions. *Mind and Language*, 17(1-2):24–36.
- Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012):1830 –1834.
- Polak, A. and Harris, P. (1999). Deception by young children following non-compliance. *Developmental Psychology*, 35(2):561–568.

- Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(04):515–526.
- Rakoczy, H., Warneken, F., and Tomasello, M. (2007). “This way!”, “No! that way!”—3-year olds know that two people can have mutually incompatible desires. *Cognitive Development*, 22(1):47–68.
- Ruffman, T., Garnham, W., Import, A., and Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? charting transitions in knowledge. *Journal of Experimental Child Psychology*, 80:201–224.
- Schneider, D., Bayliss, A. P., Becker, S. I., and Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others’ mental states. *Journal of Experimental Psychology: General*, 141(3):433–438.
- Schneider, D., Nott, Z. E., and Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, 133(1):43–47.
- Scott, R. M., Richman, J. C., and Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, 82:32–56.
- van der Wel, R. P. R. D., Sebanz, N., and Knoblich, G. (2014). Do people automatically track others’ beliefs? evidence from a continuous measure. *Cognition*, 130(1):128–133.
- Wellman, H., Cross, D., and Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, 72(3):655–684.
- Wimmer, H. and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, 13:103–128.