

How do humans develop two systems for mindreading?

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

Abstract

Experimental evidence, and reflection on everyday examples suggest that adults' mindreading is both flexible and sophisticated (as when judging someone's guilt in a court of law) and fast and efficient (as when taking part in a competitive sport). However, conceptual analysis, and evidence from analogous cognitive domains, suggests that a single cognitive process for "theory of mind" is unlikely to be able to meet these diverse demands. This leads to the proposition that adults have "two systems" for mindreading that make different processing trade-offs between flexibility and efficiency (Apperly & Butterfill, 2009).

How are adults' two systems related to the mindreading abilities observed in infants? One possibility is that the infant system grows up: although simple and cognitively efficient at the outset the infant system becomes increasingly sophisticated as children gain conceptual, linguistic and executive resources. On this view, the cognitively efficient abilities of adults must have some other origin, perhaps in automatization of abilities that were previously effortful. A second possibility is that the infant system persists: the system that explains efficient mindreading in infants also explains efficient

mindreading in adults. On this view the flexible and sophisticated abilities of adults must have some other origin, perhaps in the protracted developments charted in traditional studies of children's "theory of mind". We will draw inspiration from Susan Carey's work on "signature limits" in other aspects of core cognition to identify evidence that might distinguish between these accounts.

1. Introduction

How do humans develop two systems for mindreading? Our question presupposes that human adults (at least) do have two systems for mindreading. Actually this is something we want to argue for rather than assume. So let me start by making a distinction and an observation ...

2. slide

The distinction is between processes which are automatic and those which are not.

Roughly speaking, a process is *automatic* if it occurs irrespective of its relevance to the particulars of an agent's motives and aims.

3. slide

The puzzle starts with a question. Is adult mindreading automatic or not?

4. slide

Some evidence suggests that it is. For example, Kovacs and colleagues (2010) asked adults to identify the location of a ball. They found that adults were significantly slower to identify the ball's location when an onlooker had a false belief about the

location of the ball, even though the onlooker's belief was not relevant to the task at all. This and related findings seem to suggest that, at least in adults, mindreading is automatic.

5. slide

By contrast, Back & Apperly (2010) found that subjects are significantly slower to answer an unexpected question about another's belief when that belief is false compared to when it is true. This suggests that, at least in adults, mindreading is not automatic.

So my question is whether adult mindreading is automatic. And this question is puzzling because there seems to be evidence for both positive and negative answers.

6. Ian's first bit

7. Steve's second bit

So the question is how we develop two systems for mindreading, one automatic the other not. Two hypotheses are:

- a. automaticity through automatization
- and

- b. original automaticity.

(Define 'originally automatic' mindreading as mindreading that is automatic and not a consequence of automatization.)

Our suggestion is that (i) automatic mindreading systems are subject to arbitrary limits (because automaticity involves trading flexibility for efficiency), and (ii) we can distinguish the two hypotheses by finding arbitrary limits on automatic mindreading in infants and adults. If the limits match, this is evidence for original automaticity. If the limits differ, this is evidence for automaticity through automatization.

What sorts of limits might we expect to find on originally automatic mindreading?

Full-blown, explicit mindreading involves representing beliefs, desires and other propositional attitudes. Propositional attitudes—including beliefs—are states which form complex causal structures, have arbitrarily nest-able contents, and are individuated by their causal and normative roles in explaining thoughts and actions. If anything should make a cognitive process effortful, it is surely representing states with that combination of properties

So there are strong theoretical reasons to suppose that originally automatic mindreading could not involve representing propositional attitudes as such.

8. slide

Elsewhere Ian and I have suggested, in effect, that originally automatic mindreading involves representing relational attitudes ...

9. slide

Paradigmatic examples of relational attitudes are things like being excited by an event, encountering an object, or wanting some type of thing.

10. slide

whereas propositional attitudes have arbitrarily nestable contents, relational attitudes have no contents

whereas propositional attitudes have uncodifiably complex effects on action, relational attitudes set parameters for action

and whereas representing relational attitudes would permit a mindreader to track false beliefs involving mistakes about appearances, identity and existence, representing relational atti-

tudes would allow one to track only a limited range of false beliefs about properties such as location, colour and shape.

So consider this hypothesis: whereas non-automatic mindreading involves representing propositional attitudes, automatic mindreading involves representing relational attitudes only.

This hypothesis, if true, would explain why automatic mindreading is less demanding than full-blown mindreading.

11. slide

This hypothesis also has the virtue of indicating what sort of limits could distinguish automatic from full-blown mindreading

...

12. slide

for the hypothesis suggests that

- automatic mindreading should be good for level-1 perspective taking tasks, but not for level-2 perspective taking tasks.

Some work on the level-1/level-2 distinction has already been published, and this does suggest that for both children and adults, automatic mindreading may be sufficient for success on level-1 but not level-2 perspective taking tasks.

But since this has already been published I don't want to cover that here.

13. slide

The hypothesis about relational attitudes also suggests that

- automatic mindreading should be good for tracking false beliefs with a simple predicate-object structure, but not for tracking false beliefs which essentially involve the non-existence of something.

14. slide

and, finally, the hypothesis suggests that

- automatic mindreading should be good for tracking false beliefs about the location of something, but not for tracking false beliefs which essentially involve mistakes about identity.

So one nice feature of this hypothesis is that it makes several independently testable predictions about the limits of automatic mindreading.

If we want to find out whether adult automatic mindreading and infant mindreading are subject to the same limits, it makes sense to start by looking at these limits.

I want to finish by focusing on the issue of false beliefs about location vs. false beliefs about identity.

What do we mean by a mistake about identity?

Here's an illustration. Suppose that a single person has two roles that don't often go together, say she is both a school teacher by day and an assassin by night. If you shot the assassin, you might be surprised that the school teacher doesn't turn up to work. They are one and the same person but you didn't know that. This would be a mistake about identity: in mistakenly expecting the school teacher to turn up, you have failed to realise that she is the assassin. Your mistake could be predicted by subjects who can represent propositional attitudes as such. But it cannot be predicted by subjects who represent relational attitudes only.

15. slide

Here's how this idea might work in a format suitable for infants and adults (without the shooting).

On the screen you can see the back of the subject's head. He is sitting opposite the protagonist. On the table there is a screen. Behind the screen there is a puppet. The puppet will serve as the school teacher-assassin. The protagonist can't see

behind the barrier. She can't see the puppet right now. But the subject can see everything.

16. slide

This is the same thing, just smaller so that I can show you how the story unfolds.

I should say that the puppet has a face and costume with two aspects. Seen from one side she looks like an assassin. Seen from the other she looks like a school teacher.

17. slide

This is the same thing, just smaller so that I can show you how the story unfolds.

I should say that the puppet has a face and costume with two aspects. Seen from one side she looks like an assassin. Seen from the other she looks like a school teacher.

18. slide

The puppet now rotates behind the barrier. The subject can see this but the protagonist can't.

19. slide

Now the puppet goes out from behind the barrier. This time she looks to the protagonist as if she is an assassin. The protagonist shoots her and she falls off the table. So to the protagonist it looks like the assassin has been shot. But the subject knows that the assassin is the school teacher.

20. slide

In the final phase, the protagonist reaches around the barrier.

If the subject can represent propositional attitudes as such this makes perfect sense: the protagonist wants to retrieve the school teacher.

But if the subject is representing relational attitudes only, this reaching doesn't make any sense. For the assassin is the school teacher, and so one's disappearance is the other's disappearance.

So the protagonist's reach should be expected to subjects who represent propositional attitudes but unexpected to subjects who represent relational attitudes only.

[Don't mention control conditions]

21. slide

So this was our idea. We want to know whether adults' automatic mindreading is (a) automatic through automatization or (b) originally automatic.

Our suggestion is that if adult automatic mindreading is originally automatic, then it should be subject to the same arbitrary limits as infant mindreading.

And there are theoretical reasons to suppose that originally automatic mindreading would involve representing relational attitudes only, not propositional attitudes.

Someone who represents relational attitudes only should have no problem tracking false beliefs about location but should fail to track false beliefs about identity.

And, as I've just tried to illustrate, it would be possible to test for abilities to track false beliefs about identity using designs very similar to those used to track false beliefs about location.

So here is a way to test for original automaticity. If we find that infant mindreading is like adult automatic mindreading in that neither enables them to track false beliefs about identity, then we have evidence for original automaticity. But if it turns

out that infants and adults differ in this respect, then we have evidence for automaticity by automatization.

22. Ian's conclusion
