In several recent papers on mindreading and belief-ascription, Ian Apperly and his colleagues have reported evidence suggesting that the process whereby human adults ascribe false beliefs to others is not automatic. They have further argued that efficiency and flexibility make competing and inconsistent demands on the ability of human adults to reason about others' beliefs. To solve this tension, they have argued for the view that there are two (not one) systems of belief-ascription: an efficient but inflexible system, shared by human infants and adults, underlies the ascription of belief-like states and a flexible but inefficient system (only present in adults) underlies the ascription of genuine beliefs. If Apperly and his colleagues are right, then this two-systems model might help solve a fundamental puzzle in the developmental psychological study of belief-ascription in human children. Are they?

The puzzle in developmental psychology

When asked to predict where an agent with a false belief about an object's location will look for it, young children who know the location of the object have been demonstrated to reliably fail until they are in their fifth year. However, on the basis of the violation-of-expectation paradigm (involving a sequence of familiarization and test trials), Onishi and Baillargeon reported in Science that 15-month-olds look longer at test trials in which an agent searches for an object in one of two boxes when either she falsely believes it to be in the other box or she knows it to be in the other box. Surian, Caldi and Sperber further reported that 13-month-olds look longer at test trials in which an agent retrieves its preferred food when it is hidden from view by a high barrier than when it is visible and when the food hidden from view by a barrier has been placed there in the agent's absence than in the agent's presence. The puzzle has recently been exacerbated by the startling finding reported by Kovacs, Téglás and Endress, and recently discussed on this blog by Davie Yoon. It suggests that infants' looking behavior was influenced by the smurf's false expectation. But if so, then 7-month-olds have the ability to track another's false belief.

Until recently there were two main strategies to deal with this puzzle. One can take the data on preverbal human infants at face value and explain why it is so hard for 3-year-olds, who know about the location of an object, to predict where an agent with a false belief will look for it. Alternatively one can deny that preverbal human infants are able to represent another's false belief and try to offer deflationary low-level explanations for the data on preverbal human infants.

First strategy: infants can represent false beliefs (but it takes more to pass the standard false belief task)

Advocates of the first strategy (such as Alan Leslie and Renée Baillargeon) must explain why 3-year-olds fail the elicited false belief task if preverbal human infants are able to represent an agent's false belief. Both elicited tasks and violation-of-expectation spontaneous tasks can probe participants' ability to represent an agent's false belief. But there are at least five basic differences between the demands imposed by both kinds of tasks, which explain why representing an agent's false belief is not sufficient for success in an elicited false belief task.

(1) Unlike a participant in an elicited task, a participant in a spontaneous task is not asked any question.

(2) A participant could not succeed in an elicited task unless she understood the language in which the question is being asked.

(3) Furthermore, she must understand that what she is being asked is neither where the agent should look, nor where the marble is, but instead: where will the agent look?

(4) Only in an elicited task, in addition to forming a representation of the agent's false belief, the participant must also understand that this representation is relevant to answering the question she is being asked.

(5) The participant must be able to inhibit any prepotent tendency to answer the test question based on her own true belief whose content conflicts with the content of the agent's false belief.

Furthermore, before Onishi and Baillargeon's report, there were a few hints that the ability to represent an agent's false beliefs might not be sufficient in order to successfully pass the standard FB task. For example, in a 1994 paper, Clements and Perner noted that children aged 2 years and 11 months, who failed the standard FB task, looked at the empty location where the agent falsely believed the object to be located.

Second strategy: infants cannot represent false beliefs (and evidence to the contrary must be explained away)

Perner and Ruffman have taken the second option and tried to dismiss the view that preverbal human infants can understand an agent's false beliefs by offering various low-level explanations of the data on preverbal human infants. Among such low-level explanatory strategies, they have considered the possibility that in the course of the familiarization trials of the Onishi and Baillargeon spontaneous task, infants form a three-way association linking agent-object-location, so that they look longer at test trials that deviate most from this association. They have also considered the possibility that infants rely on simple behavioral search heuristics. In a 2007 paper, Southgate, Senju and Csibra have further considered the possibility that, far from ascribing false beliefs about an object's location, infants (i) take the agent to be ignorant of the object's location and (ii) expect an ignorant agent to search for an object in the incorrect location.

There is evidence against each of the low-level hypotheses. I start with the three-way association linking agent-object-location.

Luo and Baillargeon report that after seeing repeatedly a box move towards a cone on its right in the presence of a cylinder on its left, and after the objects have exchanged their spatial positions, 5-month-olds look longer when the box moves to the cylinder (now located where the cone was in the familiarization condition) than when the box moves to the cone (now located where the cylinder was in the familiarization condition). It is plausible to interpret this finding as evidence that 5-month-olds ascribe to the box a preference for the cone over the cylinder. Now, the three-way association linking agent-object-location hypothesis faces the following question: which deviates most from the three-way association — a change of toy or a change of location? How can one tell? Furthermore, in a single-object condition in which the cylinder was missing in the familiarization trials and the infants repeatedly saw the box move towards the cone, Luo and Baillargeon report that in the test trials infants look equally whether they see the box move to the cone at a new position or to the cylinder at the position previously occupied by the cone. Arguably, in the single-object condition, infants lack evidence for ascribing a preference to the box. But given that infants formed the very same three-way association in both the two-objects and the single-object conditions, the three-way association predicts that infants should respond the same way to the test trials in both conditions. But they do not.

According to Perner and Ruffman's search rule heuristic, an agent who is ignorant of an object's current location will search for it where she last saw it. As Surian and colleagues have observed, this looks like an ad hoc generalization that faces the following dilemma: either preverbal human infants have innate knowledge of this rule or they learnt it through experience. To assume the former is to make a strong and prima facie arbitrary nativist commitment. To assume the latter is to open oneself to the question: given that people commonly pick up an object in plain sight or look for it in several different places before finding it, on what evidence could human infants acquire knowledge of this rule? Furthermore, Baillargeon and colleagues have reported findings in which infants are presented with an agent whose beliefs do not guide her search behavior.

Finally, preverbal human infants have been assumed to expect ignorance to lead to error: if an object is moved from one location to another in the absence of an agent, infants expect the agent to search in the incorrect (initial) location. However, Scott and Baillargeon report that 18-month-olds look longer when an agent with a false belief retrieves an object at the right location than at the wrong location, but they look equally at the two events when the agent is ignorant about the location of the object.

Third strategy: a dual-systems account

The puzzle has recently been revisited by Ian Apperly and colleagues, who argue for yet a third strategy, based on a two-systems (or dual) approach to belief ascriptions. So far few experiments have been run to test the ability of healthy human adults to perform tasks of belief-ascription. In a reaction-time experiment reported by Apperly, Riggs, Simpson, Chiavarino and Samson, participants saw video stimuli in which a male actor hid an object in one of two boxes in the presence of a female actor, who looked at the boxes and indicated where the object was. After she left the room, the man swapped the locations of the boxes. Then the video paused and participants were requested to answer one of two sentence probes, one about the location of the object, the other about the woman's belief about the location of the object. After the participants responded to the probe, the video resumed until they were requested to point to the final location of the object. In the so-called "incidental false belief task", before seeing the video, participants were merely instructed to identify the location of the object at the end of the trial. In a second condition, participants were asked to keep track of both the woman's beliefs and the location of the object, while in the third condition, they were asked to keep track of the woman's beliefs only. Apperly et al. report that while participants were significantly slower at answering the unexpected belief probe than the reality probe in the incidental false belief condition, this difference disappeared in the other two conditions. They interpret this result as evidence that human adults do not automatically infer false beliefs when faced with a stimulus that affords a belief inference. They conclude that belief-ascription by human adults is not an automatic process but is sensitive to the relevance of belief reasoning for the on-going task and to explicit instructions to infer the contents of others' beliefs. They further argue that belief reasoning in adults "makes relatively high demands on scarce cognitive resources for working memory and executive control".

In accordance with the evidence suggesting that preverbal human infants can represent an agent's false beliefs, the speed and manifest lack of effort in everyday communication further suggest that the necessary theory-of-mind computations must likewise be made quickly and efficiently by human adults. If so, then human infants and human adults alike need a fast and efficient system for computing others' false beliefs. At the core of the two-systems (or dual) approach to belief-ascription recommended by Apperly and Butterfill is the distinction between two cognitive systems, one of which is efficient and inflexible and the other of which is flexible and inefficient because it makes higher cognitive demands on memory and executive functions. The first efficient and inflexible system, shared by human infants and adults, is a capacity for ascribing, not genuine beliefs, but belief-like states, which they call registrations.

The challenge faced by the dual account of belief-ascription is whether registrations differ enough from genuine beliefs for the efficient and inflexible system to differ from the flexible but inefficient one. I can see two reasons to doubt it.

First of all, Apperly and Butterfill's notion of an encountering seems to face a dilemma: on the one hand, it is supposed to be an extensional ternary relation (between an individual, an object and a location) such that an individual cannot encounter an occluded object. But on the other hand, Apperly and Butterfill are willing to relax the condition on occlusion so that objects that are "either noisy or moving on a natural trajectory" can be encountered. But if a noisy or a moving object can be encountered, then encountering cannot be defined as an extensional ternary relation between an individual, an object and a fixed location.

Secondly, an individual can be a constituent of the registration relation even when the underlying encountering relation that held at t fails to obtain at t + 1, because either the individual or the object has moved since t. If so, then the individual that is part of the registration relation must be endowed with suitable memory resources enabling him or her to mentally represent the object at a location where it no longer is. But to ascribe to an agent a false registration of an encountering relation (between the agent, an object and a location) that no longer obtains just is to ascribe to the agent a false belief about the object's location. It follows that there is no genuine distinction between the efficient and inflexible and the flexible and inefficient systems of belief-ascription. But if so, then we are left with two strategies, not three, for solving the puzzle of belief-ascription.

Key references

(for a list of Apperly's downloadable papers, click here.)

Apperly, I.A., Riggs, K.J., Simpson, A., Samson, D., & Chiavarino, C. (2006). Is belief reasoning automatic? Psychological Science, 17(10), 841-844.

Apperly, I.A. & Butterfill, S.A. (2009) Do humans have two systems to track beliefs and belief-like states? Psychological Review, 116(4), 953-970.

Back, E. & Apperly, I.A. (2010) Two sources of evidence on the non-automaticity of true and false belief ascription. Cognition, 115(1), 54-70.

Baillargeon, R., Scott, R.M. and He, Z. (2010) False-belief understanding in infants. Trends in Cognitive Sciences, 14, 3. (here)