Do chimpanzees reason about beliefs?
Kristin Andrews

1. Introduction.
For almost forty years psychologists and philosophers have been devising experiments and testing chimpanzees on the question first asked by Premack and Woodruff in 1978: “Does the chimpanzee have a theory of mind?” With this question, they meant to ask whether chimpanzees do what they assumed we do, namely attribute beliefs and desires in order to predict behavior. This capacity is known as mindreading or theory of mind. After thirty years of chimpanzees failing all the tests we put to them, some researchers concluded that chimpanzees probably don’t reason about belief (Call and Tomasello 2008). ¹ Ten years later, those same researchers were part of a team that reversed course: “our results, in concert with existing data, suggest that apes solved the task by ascribing a false belief to the actor, challenging the view that the ability to attribute reality-incongruent mental states is specific to humans” (Krupenye et al. 2016). ²

The claim that passing the false belief task is evidence of false belief ascription is one that requires critical scrutiny. For one, there is no consensus on what is involved in ascribing belief, given the lack of agreement regarding the nature of belief. And, given that we don’t directly observe anyone ascribing belief, but infer it from behavior, we must consider alternative explanations for the behavior. However, the typical alternative explanations considered are all of the variety that apes are not mentalists. There is good reason to think that apes are mentalists who see other apes and other animals as intentional agents. Nonetheless, they may not be

---

¹ At the end of their review of the status of chimpanzee theory of mind research program after 30 years of research, Call and Tomasello wrote, “chimpanzees probably do not understand others in terms of a fully human-like belief–desire psychology in which they appreciate that others have mental representations of the world that drive their actions even when those do not correspond to reality. And so in a more narrow definition of theory of mind as an understanding of false beliefs, the answer to Premack and Woodruff’s question might be no, they do not. Why chimpanzees do not seem to understand false beliefs in particular – or if there might be some situations in which they do understand false beliefs – are topics of ongoing research” (Call and Tomasello 2008: 191).

² An important lesson to be taken away from the 40 years of testing chimpanzees on false belief tracking is that we need to be very cautious making claims of inability; not finding evidence of a capacity in an experimental setting can say more about the researchers than the chimpanzees. It took people who understand the chimpanzee’s point of view--who think like a chimpanzee--to create materials that would interest a chimpanzee. The development of eye-tracking technology was also crucial.
mentalists who ascribe beliefs.

I will argue that before accepting that apes ascribe false beliefs we need to specify what is involved in ascribing beliefs, and we need to consider other mentalist hypotheses for how apes might pass. This doesn’t mean that chimpanzees don’t understand false beliefs in much the way we do. It might be instead that we understand false beliefs in much the way chimpanzees do.

2. Forty years of false belief tests

Premack and Woodruff asked whether a 14-year-old chimpanzee named Sarah could attribute states of mind to a human actor in order to predict what the actor would do next. They asked Sarah to watch videos of a human trying to achieve a goal, such as warming up a cold room with a heater or opening a door. Sarah was asked to indicate the human’s goal by choosing a photograph of the object the human would need to achieve their goal, e.g., a key to open the door. Sarah solved these problems (except for human actors she didn’t like; she tended to choose a useless object when her enemy was portrayed in the video). Premack and Woodruff concluded that Sarah understood that the human actor had a certain intention, and that she ascribed intention and knowledge (or desire and belief) to the actor.

Critics were not convinced. Sarah could have solved this task using a simple form of associative reasoning, by associating the match with the heater, and the key with the door, without having to think anything about the mental states of the actor. To avoid this simple associative explanation for Sarah’s behavior, Daniel Dennett (1978), Jonathan Bennett (1978), and Gilbert Harman (1978) each suggested an alternative experiment based on asking whether chimpanzees can think that others have false beliefs. If a person falsely believes that two things go together, then simple association will not allow a subject to predict how the person will act, because the two things in fact do not go together. Instead, to predict a falsely-believing person’s action a subject would have to consider how the person sees the world. If we accept that belief and desire together cause behavior, and that beliefs and desires are propositional attitudes—representational states consisting of an attitude toward a proposition which can be true or false—then to predict someone’s false belief behavior one could attribute the false belief along with a desire to the target. For example, if we can say that Sharine believes that the key will open the door, and that she desires to open the door, we can predict that Sharine will insert the key into the door’s lock, even if we know that the key doesn’t fit the lock.

This idea led to the moved object false belief task, which became the standard means for testing belief reasoning in children (Wimmer and Perner 1983). A version goes like this: children watch a puppet show in which Maxi hides a piece of chocolate and then leaves the room. While Maxi is out, his mother finds the chocolate and moves it to another location. Then Maxi comes back to get his chocolate. The story is stopped and children are asked where Maxi will go to look for his chocolate. Children who predict that Maxi will look for the chocolate where he left it pass the test. Children who predict that Maxi will look for the chocolate where it really is fail. Passing the test is interpreted as being able to reason about beliefs.

While human children pass this task between 4 and 5 years old (Wellman et al. 2001), and human infants appear to pass a nonverbal version of the task in the first two years of life (e.g. Buttelmann et al. 2015; Onishi and Baillargeon 2005; Southgate et al. 2007), chimpanzees failed many versions of this task given over the years. In one, chimpanzees learned that a communicator could see in which of two boxes a hider placed a food reward, though the
chimpanzees could not see for themselves (Call and Tomasello 1999). After the hider placed the food, the communicator marked the box that contained the treat by placing a token on top of the box. Then the chimpanzees had the chance to choose one box and receive the contents. Once chimpanzees were competent at this task, they were given the false belief task. The situation started out the same, but after the hider placed the treat, the communicator left the room; and while she was gone, the hider switched the location of the two boxes, which were identical in appearance. When the communicator returned, she marked the box sitting in the location of the original placement of the treat, which, unbeknownst to her, was empty. Chimpanzees, like four year-old children, failed this test. Five year-old children, however, passed. Despite efforts to change the structure of the task by making the chimpanzee compete with a human (Krachun et al. 2009) or with another chimpanzee (Kaminski et al. 2008), and by removing both cooperative and competitive aspects (Krachun et al. 2010), these changes failed to elicit false belief tracking.

What all these attempts had in common was using food to motivate chimpanzees to track false belief behavior. The experimental material that finally motivated chimpanzee to pay attention to false belief had nothing to do with food (Krupeyne et al. 2016). The story the chimpanzees watched involved a human who was attacked by someone in a gorilla suit (King Kong). Chimpanzees first saw King Kong attack the human, and then run into one of two haystacks to hide. The human grabbed a stick and hit the haystack where King Kong was hiding. In the false belief conditions, the human had to leave the scene to get the stick, at which point King Kong changed position and left the scene. When the humans came back, the chimpanzee subjects looked at the haystack in which King Kong hid before the human left, anticipating that the human would beat that haystack with the stick (See Figure 1).

For a detailed description of these studies, see Andrews (2017).
One innovation which made this experiment successful was using chimpanzee-relevant materials. Another was the eye-tracking technology and chimpanzee training which permitted careful analysis of where chimpanzees were looking. In addition, this study was modeled on the study widely touted as demonstrating belief reasoning in 2-year-old children, including all the controls and the test trial types (Southgate et al. 2007).

The authors think that this study shows that chimpanzees can ascribe false beliefs. This parallels Southgate and colleague’s interpretation of the toddler study: “The data presented in this article strongly suggest that 25-month-old infants correctly attribute a false belief to another person and anticipate that person’s behavior in accord with this false belief” (Southgate et al. 2007, 590).

However, we should only conclude that infants and apes attribute beliefs to others if we first understand what is involved in attributing beliefs, and if attributing beliefs is the best explanation for passing this task.

3. What is involved in attributing beliefs

To understand belief mindreading, we can start by comparing it with other kinds of mindreading. Over the past 15 years we’ve seen a number of studies that appear to demonstrate that chimpanzees mindread perceptions (see Butterfill, Clatterbuck, Halina, Lurz this volume). Perceptual mindreading, broadly defined, is the capacity to understand what others can and cannot see. Sub-types of perceptual mindreading can also be identified. According to Flavel (1974), level-1 perspective taking consists of knowing that what one sees may differ from what others see, whereas level-2 perspective consists of knowing that how one sees things can differ from how others see things. While children have been tested on both level-1 and level-2 perspective taking, the chimpanzee perceptual mindreading studies have been focused only on level-1 perspective taking.

Level-1 perspective taking differs from both level-2 perspective taking and mindreading belief in that level-1 perspective taking is conceptually light, level-2 perspective taking is a bit heavier, and mindreading belief is heavier yet. To explain, let’s compare the three kinds of mindreading:

(A) Mindreading level-1 visual perception: Mindreader believes that subject sees object.

(B) Mindreading level-2 visual perception: Mindreader believes that subject sees object as q.

(C) Mindreading belief: Mindreader believes that subject believes that P.

Now let’s consider the cognitive and conceptual requirements for each of these types of mindreading. They all require that the mindreader has beliefs, and that is a point granted by all parties engaged in the debate. For (A), a mindreader only needs the concept of see. The object can be understood as “this thing” de re and need not be characterized in terms of how it appears to another, de dicto. What is it to understand that another can see a thing? That’s a good question. Lurz (this volume) suggests that seeing is a mental state concept, and it is distinct from what he calls “direct line of gaze” -- an observable relational concept between the body of a gazer and an object. He has previously identified the mental state concept of seeing with level-2
perspective taking (Lurz 2011), and presumably, given that it is a mental state, at least part of what would be needed to know that another sees something as something else is, plausibly, some understanding of the other as a conscious intentional agent.

Once we get to (B), we see that the mindreader needs the concept see along with additional, and contrasting, descriptive concepts. The level-2 perspective taker can know that the eraser looks like a piece of candy to someone else. This adds a greater conceptual competence to the mindreader, and requires the ability to consider the mode of presentation of the object to another person, and to understand that a perspective might not correspond to reality.

In (C), a belief mindreader has an even greater conceptual burden, since one has to have the concept believe as well as the ability to correctly attribute propositional attitudes. What is belief? That’s a really good question. Belief is typically understood as an attitude toward a truth-evaluable proposition which is mentally represented, though the vehicle of the mental representation is debated (see Bermúdez, Glock, Rescorla this volume). We might represent propositions sententially in a language of thought, as maps, as causal relations, as probabilistic relations, or as some combination of these. Of course, the cognitive load of belief attribution will differ depending on the vehicle of belief in play.

In addition to those who disagree about the vehicle of belief, there are others who do not accept the representational nature of belief. Dispositionalists or instrumentalists such as Braithwaite (1933), Davidson (1984), Dennett (1987, 1991, 2009), Marcus (1990, 1995), Ryle (1949), Schwitzgebel (2001, 2002) and Sellars (1981) all advocate various alternative accounts of belief. (See also Vincent and Gallagher, this volume). So-called radicals in cognitive science deny any role for representation, given their view that explanations in terms of dynamical systems can explain our cognitive capacities (e.g. Barrett 2011; Chemero 2009; Hutto and Myin 2012). Yet others think that much of what has been called mental representation can be understood in terms of externalized deeds or action (e.g. Rowlands 2006; Thompson 2007; Varela et al. 1991).

Finally, the folk have their own understanding of belief, though, English speakers in North America typically use the word “think” rather than “believe” (Buckwalter et al. 2013). In the colloquial understanding, thinking that P can either mean that one represents and stores P as information, or in a thicker sense, it can mean that one might “like it that P is true, emotionally endorse the truth of P, explicitly avow or assent to the truth of P, or actively promote an agenda that makes sense given P” (Buckwalter et al. 2013, 2). The thicker sense of “think” is consistent with representational, dispositional, and enactive views of belief.

So, what it means to have a concept of belief is disputed, and different theorists will have different criteria for thinking that one has a belief. Take three quick examples. On Schwitzgebel’s phenomenal-dispositional account of belief, “To believe that P, on the view I am proposing, is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P. What respects and degrees of match are to count as 'appropriate' will vary contextually and so must be left to the ascriber's judgment” (Schwitzgebel 2002, 253). The dispositional stereotype for a belief that P will consist of the cluster of behavioral and phenomenal dispositions we tend to associate with the belief. And the dispositional stereotype for a belief that someone has a belief would likewise consist of the cluster of behavioral and phenomenal dispositions associated with thinking that someone else has the set of behavioral and phenomenal dispositions associated with the attributed belief. So,
for example, to believe that a person believes that there is a monster in the rightmost haystack may be associated with the disposition to expect that the person feels more scared approaching the rightmost haystack, runs away from the rightmost haystack, attacks the rightmost haystack, etc., and to be surprised if the person didn’t act as expected. The Krupenye study would suggest the chimpanzee has this belief about the human actor’s belief, on the phenomenal dispositional account.

On Dennett’s intentional systems theory, beliefs are interpretive gambits that permit us to better understand and predict patterns of behavior. Because believers are rational, they act in predictable ways that are explained in terms of how they should act given the beliefs they should have. This intentional system is holistic, such that the mental concepts of the system are related to one another, and to behavior, providing a coherent explanatory system. As a holistic system, it is important to interpret behaviors from within a larger context, because the interpretation of a single behavior is a function of how it fits into the larger pattern of behavior.

If the intentional system includes the platitude that seeing is believing, then, given the evidence across circumstances that chimpanzees understand seeing, the Krupenye findings would support interpreting all the mindreading studies as evidence of mindreading belief. If it doesn’t, then we can interpret the Krupenye findings as evidence that chimpanzees understand seeing, but not belief, because the chimpanzees could have anticipated the human’s action because that’s where the human last saw King Kong, and people seek out things where they last saw them. Either way, the Dennettian approach requires us to take into account a large body of behaviors in order to know how best to interpret any one of them, and it would be a mistake to interpret the result of a single study outside of the larger context.

Because an intentional system tracks robust patterns, an interpretation would also have to take into account the chimpanzees’ failure to track false belief in previous studies. Whether the pattern is robust enough to warrant the interpretation of belief attribution would be open to debate, and likely would require further evidence. The intentional systems view may be incompatible with Krupenye and colleagues’ explanation for why apes failed earlier false belief tests: “Differential performance between tasks may reflect differences in task demands or context, or less flexible abilities in apes compared with humans” (Krupenye et al. 2016, 113). If false belief capacities in chimpanzees are much less flexible than those in humans, such that the pattern doesn’t hold, then on the intentional systems view we shouldn’t call them false belief capacities.

Like Dennett’s intentional systems theory, on Davidson’s interpretationism a belief is also understood in terms of attributions, but on this account the attributions are essentially tied up to language. We understand other creatures by applying a principle of charity to them, thinking that they are rational and that their utterances are largely true. Furthermore, we understand other creatures by thinking that they understand us in much the same way. Davidson writes, “My thesis is rather that a creature cannot have a thought unless it has language. In order to be a thinking, rational creature, the creature must be able to express many thoughts, and above all, be able to interpret the speech and thoughts of others” (1982, p. 100). Thus we see that the only empirical data that would be relevant to determining whether a chimpanzee could mindread belief would come from a research program studying linguistic abilities.

And finally, consider one realist account of belief, Fodor’s Language of Thought (LOT). On this account, a belief is a representation of sentential content in a language-like vehicle (Fodor
1975; see Beck this volume). On this view, in order to mindread beliefs one must be able to represent a belief about someone else’s belief, which requires having the concept of belief. Like on Davidson’s interpretationism, a LOT mindreader needs a significant amount of conceptual resources, including the concept of belief, in addition to linguistic competence (though unlike Davidson, Fodor thinks that having an external language isn’t required for thinking in the language of Mentalese). That is, a mindreader would have to represent mental sentences about others’ mental sentences, attribute those mental sentences, and understand that the other is thinking the target mental sentence. On this view, passing the false belief task wouldn’t automatically provide evidence of belief ascription, if alternative LOT explanations are possible. Insofar as there are alternative mentalist explanations for passing the task, which I argue for in the next section, passing wouldn’t entail false belief ascription on the LOT account of belief, either.

The take away message is that we can’t say that chimpanzees or infants attribute beliefs without specifying the account of belief that we are working with. These and other accounts of belief, rationality, and representation (see chapters by Beck, Bermúdez, Boyle, Gauker, Proust, and Rescorla) can also provide different criteria different explanations for what passing the false belief task requires. They differ in how intellectual belief ascription is, yet, they all can explain passing the tasks in terms of attributing beliefs.

Let us now look briefly at explanations for passing false belief tasks that do not require belief attribution.

4. Alternative mentalist explanations for infant and chimpanzee false belief prediction

Is there a way to be a mentalist and predict false belief behavior without being a belief mindreader? There are three main approaches: A two-systems account, a developing one-system account, and a pluralistic account.

Apperly and Butterfill (2009, 2013) defend a two-system account of mindreading, according to which there is a fast automatic system that permits tracking false beliefs without representing them as such, and a slower conceptual system that takes the familiar form (see Butterfill, this volume). Key to the view is how the early-developing system trades in non-psychological proxies for mental states: “encountering” and “registration” (Apperly & Butterfill 2009 p. 962). The encountering representation tracks perceptions, and so would rule out objects being in a target’s field if, for example, they are located behind an opaque barrier relative to the target. Registering an object involves encountering it at one location and nowhere else since. Together with the capacity for representing goals nonmentalistically, these states allow subjects to track at least some of the mental states of targets without the conceptual apparatus required for representing belief as such. This theory would explain the ability of the chimpanzee to predict the behavior of the human who attacks King Kong as a case of the human registering King Kong in the rightmost haystack.

Carruthers advocates a developing one-system model such that from infancy humans already have concepts such as thinks, likes, is aware of, and they use these concepts in ascribing propositional attitudes to others (Carruthers 2015). This ability matures over time, as humans learn how to apply the concepts more generally by coming to see the kinds of perceptual access that give rise to various propositional thoughts. On this view, we may see the ape ability to pass this false belief task as evidence that apes have these proto-belief concepts, without having the
concept of belief, given that they work well enough to predict false belief behavior. However, unlike the children, apes would be stuck at this intellectual stage of development.

A third alternative, which I advocate, arises from a challenge to the commitment about the function of propositional attitude attributions. On both hypotheses discussed above, adult humans are thought to attribute beliefs and desires to predict behavior. While it may seem intuitive that we predict behavior by thinking about people’s beliefs and desires, that intuition might be a post hoc rationalization of our capacity to easily track behavior. The last few decades of social psychology research has taught us that humans are particularly bad at introspecting the mechanisms we use to act (e.g. Nisbett and Wilson 1977).

I argue for Pluralistic Folk Psychology (PFP), according to which we predict behavior in various ways, including: stereotypes; self-reference; primary intersubjectivity; situation; inductive generalizations over past behavior; norms; non-propositional mental states such as moods, emotions, and goals; teleology; and trait attribution (Andrews 2012; 2015a). According to Pluralistic Folk Psychology, we build models of individuals and types of individuals that consist of information such as personality traits, social roles, emotions, histories, goals (and yes, in many humans, beliefs too), and manipulate these models to predict behavior. And, even when we are not attributing beliefs (which is most of the time), we do see others as minded, intentional creatures.

Essential to PFP is the idea that prediction and explanation are not symmetrical; a model can predict without explaining. If this is right, then it is also right that the false belief moved object task will not paradigmatically elicit reasoning about beliefs. Given that we explain our own behavior in terms of our beliefs when we are looking to justify our actions (Malle et al. 2000), and that justifying anomalous behavior permits the development of cumulative culture, I argue that the function of belief attribution is to explain—and justify—anomalous behavior (Andrews 2012).

Given this brief description of PFP, we can interpret the chimpanzee false belief tracking behavior in terms of registration as Apperly and Butterfill suggest, without accepting their complete two systems account of belief. Rather, the encountering and registration relationships can be included as important parts of the folk psychological toolkit. On PFP, evidence of belief attribution would come not from passing some other form of the false belief task. Rather, it would come from observing a range of explanation-seeking behaviors, best interpreted as an attempt to understand why individuals act in the way they do, and accepting others even when they act eccentrically.

5. Conclusion

Apes and infants are sensitive to others’ false beliefs. But we can’t yet say tout court that they have a concept of belief—because we’re not yet in agreement about what counts as understanding belief, and there are alternative explanation for their actions that do not require belief ascription. But does this mean that apes lack our theory of mind?

In a commentary in Science on the Krupenye et al. study, Frans de Wall writes:

Theory of mind is probably part of a much larger picture that includes empathy, social connectedness, and the way bodies relate to other bodies. It is no accident that the tests conducted here focus on the body, i.e., subjects' eyes following the physical movements
of actors. As such, the study by Krupenye et al. may help us move away from the prevailing assumption that theory of mind relies on a cognitive simulation of what is going on in the heads of others. Reading others’ minds is beyond anybody’s capacity. All we can do—and what apes apparently do in similar ways—is read bodies (de Waal 2016: 40).

Here, de Waal challenges what has for so long been taken for granted— that we humans read minds. While we do interpret people, and offer explanations for people’s actions, we shouldn’t expect a great deal of accuracy in our explanations of others’ actions. Yet, we’re really good at coordinating our behavior with others, which means we are good at anticipating what others will do. Predicting behavior, and explaining behavior, are two different practices, and have two different functions; we need to stop seeing them as always caused by the same processes.

Why is it, then, that infants pass the implicit tasks but not until 4 do children pass the explicit verbal tasks? PFP has an answer: the verbal tests raise a puzzle for subjects, and elicit reasoning systems that need not come on line when one is simply anticipating behavior in a real life situation. The question “Where is Maxi going to look for his chocolate?” raises a challenge for children that they have to figure out. An important difference between the implicit tasks, including the Buttelmann et al. (2009) active helping task, on the one hand, and the Maxi task on the other, is that the first task is about acting, and the second is about engaging in an act of charity to figure out what a reasonable person would do in a situation like that. The explicit tasks are akin to seeking a justification.

To conclude that apes do engage in belief reasoning is premature, but, as well, it would be premature to conclude that they do not. First, we need to get clearer on what we mean by belief attribution. Second, given that it took humans 40 years of looking to find evidence that chimpanzees can track false belief behavior, we should not be too quick to judge inability, especially since we haven’t even started looking for evidence that apes explain behavior in terms of discrepant representational states. For all this time, we’ve been looking for evidence of ape belief reasoning in all the wrong places (Andrews 2005). Maybe now, with the knowledge that apes can make predictions that track someone’s false belief, we can begin to examine other belief-relevant behaviors, and then, with a particular theory of belief in hand, we might be able to say whether apes actually have the concept of belief, and whether they use it to explain the strange things their friends and families sometimes do.

Further Reading: For some alternative ways of thinking about the function of belief attribution, see Tadeusz Zawidzki’s book Mindshaping (MIT Press 2013) and Victoria McGeer’s articles “The regulative dimension of folk psychology” (2007, in D. D. Hutto & M. Ratcliffe (Eds.), Folk Psychology Re-Assessed (pp. 137–156). Dordrecht, The Netherlands: Springer) and “Mind-making practices: The social infrastructure of self-knowing agency and responsibility” (2015, Philosophical Explorations, 18(2), 259–281), as well as my book Do Apes Read Minds?, which is more about mindreading in the human ape than the nonhuman ones.

Acknowledgments: Thanks to Jacob Beck, Cameron Buckner, Brian Huss, and Brandon Tinklenberg for helpful comments on earlier drafts of this chapter.
Bibliography


