

Notre Dame Philosophical Reviews

2013.04.09

Author 

Kristin Andrews

Do Apes Read Minds?: Toward a New Folk Psychology

Published: April 07, 2013

Kristin Andrews, *Do Apes Read Minds?: Toward a New Folk Psychology*, MIT Press, 2012, 312pp., \$38.00 (hbk), ISBN 9780262017558.

Reviewed by Neil Van Leeuwen, Georgia State University and University of Johannesburg

There's no doubt it's an important problem.

Here's the familiar progression. As the ice of behaviorism melted in the second half of the 20th century, we noticed two things. First, people have minds. Second, people *think* about other people's minds. These ideas were seeds in the cognitive revolution's spring. We can study minds the way physicists study their unobservables -- for them electrons, for us beliefs -- inferring hidden causes behind data we collect. And perhaps, while being interesting in its own right, peoples' ways of thinking about other minds could give initial purchase on what psychological states are. Thus, from this second seed arose the theory of mind/mindreading/folk psychology projects that have flourished since the 1980s, with vigorous debates over whether humans understand other minds by way of theories, simulations, or -- altogether now! -- a *hybrid* of theory and simulation. The data nourishing these projects came largely from two anthropocentric sources: the false belief test, regarded as almost definitive of mentalistic understanding, and autistic humans, whose deficits highlight by contrast mindreading capacities "normal" people have.

But the frost of behaviorism has departed more slowly from our primate relatives. Though most cognitive scientists happily accord minds to non-human great apes and other animals, resistance is more common and vigorous than in the human case. "How can a creature without *language* have *beliefs*!" cries the neo-Davidsonian. More to the present point, almost up to this past decade something like behaviorism was the prevailing picture (with notable exceptions) of how non-human apes understand *other* creatures. A mere chimpanzee, most thought, couldn't do something as sophisticated as attribute unobservable mental states, like perception, belief, desire, and knowledge; they must think of other creatures only in behaviors. Every chimp was a Skinnerian.

But even that patch of frost has started to melt, which brings us to our problem and to Kristin Andrews' engaging new book. It's become clear that other great apes, like chimps, have much in common with us when it comes to doing folk psychology. They track goals and perceptual awareness in other creatures, including awareness of what *has been seen* and not just *what currently is seen* (see Call and Tomasello 2008 for a concise review). But on traditional views of mindreading in humans, to be a folk psychologist at all entails understanding *beliefs* -- especially false beliefs. The most well-designed experiments, however, don't support the view that chimps understand false beliefs or perhaps beliefs at all. So we're in a quandary: non-human great apes cognize more of genuine psychology than a brute Skinnerian could stomach, but they lack the key form of mental state attribution that was thought to constitute having a folk psychology at all. So how does their folk psychologizing work?

Andrews proposes a way out: Pluralistic Folk Psychology. Reviewing ample evidence from social psychology, she argues it is a mistake to regard propositional attitude attribution as the sole or even primary means by which *humans* predict and explain

behavior. To this end, Andrews argues that folk psychological prediction and explanation are not just mirror images: the "psychological symmetry thesis" holds that the processes of prediction are the same as those of explanation, just in reverse, and she rejects this view as a relic of Hempelian philosophy of science. Having seen this point, we can hold that we *predict* behavior by a plurality of strategies, while attributing beliefs and desires mainly to *explain* behaviors, especially those outside the norm. If her plurality thesis is true, we can then hold that non-human apes share some of our predictive folk psychological strategies, even if they lack the explanatory machinery of propositional attitude attribution. Implicitly held views about symmetry between prediction and explanation were an invisible cage, preventing theorists from realizing that humans and other primates predict behavior by means *other* than those by which humans tend to explain behavior.

This diagnosis of underlying intellectual currents may or may not be correct. But a pluralistic folk psychology is worth pursuing anyway, despite a lurking problem. So let's review some of Andrews' key theses.

Andrews discusses "four methods of predicting behavior that arise from research in social psychology: predicting from the situation, predicting from self, predicting from trait attribution, and predicting from stereotype" (p. 68). People use these regularly, she holds, *without positing beliefs or desires*. Thus, these methods could bridge human and non-human ape folk psychology.

Andrews gives the following examples of predicting from situation:

Whether someone just finished a long bike ride or disembarked from a plane in a foreign country, it is easy to generate scenarios the individual might follow. The cyclist will probably drink some water and take a shower. The passenger might rush toward the customs line or use the restroom. If you needed to find the cyclist after he finished his ride, you would look in the kitchen and the bathroom. (p. 71)

She thinks that situational prediction on the part of folk psychologists often won't involve positing beliefs or desires. Why? She argues as follows, as best I can reconstruct: working out an agent's future behavior from propositional attitude attribution often yields inaccurate predictions, while predictions from situations can be quite accurate, so it's likely that the latter prediction is typically devoid of belief and desire attribution. We shall return to this point.

Predicting from self, on Andrews' view, also doesn't involve belief and desire attribution, and in this it is cognitively simpler than simulation. If I'm predicting your behavior from myself, I needn't adopt offline versions of your beliefs or desires to run through my cognitive machinery; I just ask what I would do or feel in your situation. And Andrews cites evidence that children at least *don't* adjust for the desires of the persons they predict: "when told a story about an actor who wants another child to be hit by a ball, subjects younger than five years are unable to suppress their own sentiments and will say that the actor is sad when the child is hit by the ball (Yuill 1984; Yuill et al. 1996)" (p. 75).

Predictions from stereotypes and traits are related to each other but not identical.

"Stereotypes are sets of properties that are associated with a particular group or kind of person" (p. 81). For example, "Children's early stereotypes of sex roles lead them to expect that boys won't skip rope and that girls won't play with trucks" (p. 82). There is an upside and a downside to folk psychological stereotypes: stereotypes are cognitively efficient in being fast, easy to use, and information rich, and predictions from stereotypes are often more accurate than predictions from other strategies, but stereotypes are the stuff of prejudice and can continue to infect judgments of others even when there's evidence they don't apply.

Traits are more specific properties, not necessarily inferred from group membership. Extroversion, introversion, rudeness, kindness, stinginess, generosity, etc. are all traits. Traits attributed by the folk psychologist, as Andrews construes them, are causally efficacious dispositional mental states; people infer them quickly from very little behavioral evidence and use them

extensively in other predictions. Again a trade-off: speed, automation, and *relative* accuracy at the cost of recalcitrant error when the trait attributed does not apply.

Critically, once again, Andrews thinks folk psychologists needn't and typically don't attribute beliefs or desires when predicting from stereotypes and traits.

And there is the lurking problem, as many present readers may have suspected. When I ascribe traits like selfishness or generosity, am I not also ascribing *desires*, like desires for the self or others to have in abundance? Correspondingly, if I ascribe rudeness on the basis of a stereotype, am I not implicitly ascribing mental states, like a bad attitude? Is not the stereotyped boy thought to *believe* jumping rope is for girls? Many cognitive scientists hold that such propositional attitude attributions are made by a fast, unconscious theory of mind system. The same point applies to predictions from the self and from the situation. When I ask how you will act by asking what I would do, am I not putting myself in the situation *as seen by you* (even if I don't imagine your further desires)? When I predict from an agent's situation, do I not track what she *knows* about the situation? I predict the agent will go to the kitchen after his bike ride, if I think he knows where it is. When I think he doesn't know, I predict he will ask or look. Either way, knowledge attribution is critical to what I predict, even if I am broadly predicting from the situation. So a traditionalist about folk psychology, especially from the theory theory camp, would say that Andrews has only identified sub-categories of folk prediction that ultimately rely on implicit attribution of propositional attitudes.

Andrews is of course aware of this response to her views and deals with it in various ways, constantly attempting to minimize the role of contentful state attribution in folk psychologizing. But I must confess that I never found her arguments relating to this point compelling. Since this point is at the heart of what it is to be a folk psychologist, it's worth reviewing some of those arguments. I'll review three and raise what I take to be the core concerns.

First, early in the book, Andrews challenges the standard interpretation of the false belief task. In this task, subjects see the following: an agent leaves/sees something in location A; that something is moved to location B when the agent is away or not looking; the subject must then indicate where she thinks the agent expects the object to be. Subjects who indicate location A are standardly interpreted as attributing false beliefs to the agent and thereby "pass" the test. It used to be thought that children couldn't pass until age four or so, but recent evidence based on violation of expectation (looking times) and anticipatory looking has infants passing at younger than age two (Onishi and Baillargeon 2005; Baillargeon, Scott, and He 2010).

Andrews rejects the standard interpretation, however. Instead, she says infants may have associations or behavioral rules, may reason from ignorance, or may engage in explicit simulation. Let me address the first of these suggestions. One behavioral rule Andrews thinks may explain success on the false belief task is that people look for objects where they left them; she makes this suggestion twice in the book (p. 33 and p. 110). The problem with this explanation, however, is that on several versions of the paradigm, the agent whose behavior the subject is observing does not "leave" the object in question anywhere; rather, an experimenter's hand moves the objects around in the agent's sight or in the agent's absence (e.g., the caterpillar experiment; see Baillargeon, Scott, and He 2010, p. 111). So Andrews mooted behavioral rule, in those cases, *couldn't* be what subjects use to anticipate the agent's behavior. The interpretation that says subjects attribute false beliefs remains the only one that (on my view) concisely accounts for the *diverse* array of false belief tasks, many of which are designed with controls that rule out the sort of explanation Andrews posits.

Second, Heidi Maibom (2003, 2007, 2009) and Peter Godfrey-Smith (2005) have proposed different versions of the view that folk psychology is like theoretical modeling in science. We predict and explain the behavior of others by constructing a model of their psychology that posits various psychological states, manipulating the model to get predictions and explanations. This view has multiple advantages over more traditional versions of the theory theory, such as the fact that

there are multiple ways of construing a model, which makes sense of the range of attitudes one can take to the metaphysical status of a posit like *belief*, despite that posit's having a common folk psychological utility. Now, Andrews is more a fan of the model version of theory than of the traditional version, but she nevertheless criticizes it.

Another worry about the model theory is that it seems not to leave room for some of the automatic processes involved in our folk psychology practices. Our unmediated responses to others' facial expressions, hormones, movements, and touches suggests an immediacy in our folk psychology that is not reflected by any exhaustive theoretical account. (p. 205)

Andrews is referring to findings such as the fact that brief touch, even if we're not conscious of it, can cause us to judge the person touching to be more trustworthy. But it's hard to see why Andrews thinks model theory "seems not to leave room" for such processes. Godfrey-Smith explicitly writes, "My view can allow that elements might be *introduced* to the folk-psychological model via a number of different paths" (p. 8, his italics). Sub-personal cognition of things like light touch and oxytocin may cause us to feel a certain way toward someone, and that feeling inclines us to include *trustworthy* (or whatever) in our folk psychological model of that person. That model is then used for further prediction and explanation. So I think Andrews' critique of model theory doesn't go through. And since model theory holds that beliefs and desires are typically components of folk psychological models, it still hasn't been made plausible that we can do folk psychology without attributing them.^[1]

Third, Andrews writes, in a discussion of cognitive resources needed to perpetrate deception:

a theory of mind offers an advantage only when predicting the behavior of someone else with a theory of mind; when predicting the behavior of others who lack a theory of mind, attributing beliefs will normally offer no additional predictive power. (p. 219)

Her idea, I take it, is that attributing beliefs is only useful for predicting organisms who can also *attribute* beliefs. If this were true, then the predictive value of belief and desire attribution would be severely restricted, and it would become correspondingly more plausible that we often get by without it. But in general, having a theory of property P is useful for predicting entities that *have* property P, not only for predicting entities that also *have a theory* of property P. A theory of mass applies to objects with mass, not just objects with a theory of mass. And the point carries over to theories of mind, as far as I can tell. Knowing my cat believes her catnip mouse is under the sofa helps me predict what she'll do when I lift up the sofa. So again, I'm unconvinced by Andrews attempt at downplaying the predictive utility of a propositional attitude-attributing theory of mind.

Let's review the dialectic. Andrews wants to show there are ways creatures can do folk psychology that don't rely on belief and desire attribution. This will allow us to understand how non-human apes and other animals could be folk psychologists. Thus, she (i) identifies manners of prediction in humans that don't *prima facie* implicate such attribution and (ii) criticizes views that posit propositional attitude attribution, as a way of suggesting folk psychologists can do without it. If (ii) were successful, that would undermine the arguments of those who want to say there is mindreading implicitly going on in the manners of prediction referenced in (i). Note further that all Andrews needs is an existential claim, not a universal one, with respect to each of her manners of prediction (and explanation) that she argues do not involve propositional attitude attribution: to serve her wider aims, it only needs to be the case that we *sometimes* attribute (for example) traits without attributing attitudes; it needn't be all the time. As long as it's *sometimes*, there are at least *some* folk psychological strategies we could share with our primate cousins. But it doesn't seem to me that (ii) is successful, as I've been arguing. If I'm right, then it remains an open question whether the predictive strategies Andrews discusses in humans in fact always involve implicit propositional attitude attribution, as some think, or not. And if that's true, it's unclear to what extent those strategies can be the bridge between

human and non-human folk psychology. The problem still lurks.

I do think that Andrews has an overall valuable and interesting *approach* for showing continuity between human and non-human folk psychology: show that there are ways of predicting and explaining behavior that humans *could* share with non-human primates, given the latter's apparent limitations on understanding false beliefs. It's just that the approach, as it stands, needs additional execution. That, of course, is normal for cognitive science.

Do Apes Read Minds? has much to recommend it that I haven't had space to discuss. In addition to engaging thoroughly with the relevant social psychology and philosophy of science, it introduces a new pluralistic framework for theorizing about folk psychological explanation (Chapters 7 and 8); it introduces a new evolutionary theory of how folk behavioral explanation came to be and contrasts this view with the famous Social Intelligence Hypothesis (Chapter 11); it reviews relevant primatology literatures, like the literature on coordinated chimpanzee hunting (Chapter 12) -- and much else. So the book makes progress and is worth reading. In the end, I did feel that spring was closer to arriving for our primate relatives.

ACKNOWLEDGEMENTS

I'd like to thank Heidi Maibom, Guilherme Sanches de Oliveira, and Shannon Spaulding for feedback on an earlier draft of this review. I'd like to thank Deena Skolnick Weisberg for an email exchange that helped clarify my view of the false belief task literature.

REFERENCES

- Baillargeon, R., Scott, R., and He, Z. (2010) "False-belief understanding in infants," *Trends in Cognitive Science* 14(3), pp. 110-118
- Call, J. and Tomasello, M. (2008) "Does the chimpanzee have a theory of mind? 30 years later," *Trends in Cognitive Science* 12, pp. 187-192
- Godfrey-Smith, P. (2005) "Folk Psychology as a Model," *Philosophers Imprint* 5(6), pp. 1-16
- Maibom, H. (2003) "The Mindreader & the Scientist," *Mind and Language* 18, pp. 296-315
- Maibom, H. (2007) "The Presence of Others," *Philosophical Studies*, 132, 161-190
- Maibom, H. (2009) "In Defense of (Model) Theory Theory," *Journal of Consciousness Studies* 16(6-8), pp. 360-378
- Onishi, K. H. and Baillargeon, R. (2005) "Do 15-month-old infants understand false beliefs?" *Science* 308, pp. 255-58
- Yuill, N. (1984) "Young children's coordination of motive and outcome in judgments of satisfaction and morality," *British Journal of Developmental Psychology* 2, pp. 73-81
- Yuill, N., Perner, J., Pearson, A., Peerbhoy, D., Van den Ende, J. (1996) "Children's changing understanding of wicked desires: From objective to subjective to moral." *British Journal of Developmental Psychology* 14, pp. 457-475

[1] For Maibom, there are actually three kinds of theoretical models that agents use to predict and explain others: behavioral, social (based on role structure and purposes), and full-blown psychological. She holds a view that seems friendly to Andrews', namely, that the first two kinds of model needn't involve internal state attribution. So the availability of this version of model theory leaves me somewhat perplexed as to why Andrews bothers to criticize model theory at all. Why she would criticize

Godfrey-Smith's version of it is clearer.

Copyright © 2013
ISSN: 1538 - 1617
College of Arts and Letters
Notre Dame, IN 46556

