CellPress

Review

# What Cognitive Representations Support Primate Theory of Mind?

Alia Martin[1,*] and Laurie R. Santos[2]

**Much recent work has examined the evolutionary origins of human mental state representations. This work has yielded strikingly consistent results: primates show a sophisticated ability to track the current and past perceptions of others, but they fail to represent the beliefs of others. We offer a new account of the nuanced performance of primates in theory of mind (ToM) tasks. We argue that primates form awareness relations tracking the aspects of reality that other agents are aware of. We contend that these awareness relations allow primates to make accurate predictions in social situations, but that this capacity falls short of our human-like representational ToM. We end by explaining how this new account makes important new empirical predictions about primate ToM.**

## Trends

Previous research suggests that nonhuman primates, unlike human adults and children, do not track the beliefs of others.

However, primates do track the current and past perceptual awareness of others when predicting their behavior or competing with them.

We argue that primates succeed in ToM tasks by representing awareness relations between agents and true (but not false) information.

In contrast to other accounts arguing that primates represent knowledge and ignorance, this awareness relations account is consistent with failures in false belief tasks and with recent findings that primates competitively keep information concealed but do not actively conceal it.

The awareness relations account makes new predictions for the performance of primates in tasks that require a full-fledged understanding of the mental state of ignorance.

## A Framework for Thinking about Cognitive Representations for ToM

Humans expect that agents will act on the basis of their unobservable mental states: their beliefs, desires, and intentions. Are humans alone in positing that others have internal mental states representing the external world? Or do nonhuman primates (hereafter primates) share our human-like representational ToM? Here, we explore these age-old (e.g., [1]) questions by reviewing recent ToM experiments in primates. We offer a new theoretical proposal for the performance of primates in ToM tasks, one that makes new empirical predictions about the origins of human mental state understanding.

The goal of ToM research is to figure out whether and how organisms make predictions about the behavior of other agents based on a model of the mental states of those agents. One way an organism could predict the behavior of other agents is by tracking particular agents (e.g., mom), particular bits of information that the organism knows about the world (e.g., the apple is on the table), and the relation between the two based on cues relevant to the mental state of the agent (e.g., mom sees the apple on the table). Such representations of the relation between an agent and information could then be used to predict or interpret the actions of the agent. For example, if I see that mom is looking at an apple that I see on the table, I can represent that mom now has a connection to this information that I know to be true. I could then use this 'awareness relation' between mom and the information I know to predict how mom will behave (e.g., mom will walk to the table to grab the apple).

Awareness relations can allow organisms to make many accurate predictions about the behavior of other agents. However, organisms sometimes interact with agents who represent different information than they themselves represent; agents will act on the basis of their own personal beliefs, which in some cases might be false. As such, an organism needs a way to represent the relation between an agent and a piece of information that the organism itself does not represent as reality. Human adults use this type of 'representational relation' all the time: we understand that other people believe information that we ourselves do not currently think is true,

[1]Victoria University of Wellington, Wellington, New Zealand
[2]Yale University, New Haven, CT, USA

*Correspondence:
alia.martin@vuw.ac.nz (A. Martin).

think about information that will happen in the future, and so on. The act of representing a relation between an agent and a piece of information that is not part of our current reality is a computational challenge. To predict behavior using representational relations rather than just awareness relations, organisms first need the capacity to conceive of states of the world that are different or 'decoupled' from their own current reality (e.g., [2]). Examples of such 'decoupled' states include hypothetical or counterfactual situations (what if the apple was on the floor, not the table), pretend contexts (let's pretend the apple is on a spaceship), or even past or future states of the world (yesterday, the apple was in the cupboard). Only an organism that succeeds in conceiving of these sorts of decoupled information states can form relations between these alternative pieces of information and particular agents. For example, to represent mom's outdated belief that the apple is in the cupboard, I would have to represent what I myself know about the world (the apple is on the table) as well as a separate decoupled counterfactual state of the world (the apple is in the cupboard) to which the agent in question (mom) has a relation (mom thinks the apple is in the cupboard). The ability to form representational relations allows an organism to predict how an agent will behave when that agent's representation of the world conflicts with the organism's own idea of reality. In this way, representational relations allow for much of the richness of adult human ToM.

## The Development of Humans' Ability to Use ToM Relations

Much of the focus on ToM research to date has explored how humans develop the ability to form representational relations. Specifically, researchers have examined how humans develop the capacity to form one common representational relation: that an agent has a false belief. In so-called 'false belief studies' [3–6], a character typically hides an object in one location and the object moves to a new location while the character is gone. The question of interest is where participants predict the character will search for the object. Successful performance on this task requires that participants do more than merely establish an awareness relation between the character and the object; participants who treat the character as unaware about the location of the object should expect the character to search at random. By contrast, participants who are able to form a representational relation between the character and the object should realize that the character is not merely unaware: the character should search for the object in its original location, where the character falsely believes the object to be. Although earlier research suggested that children do not begin to form representational relations until 4 years of age [3], recent work suggests that human infants show some evidence of using representational relations within the first 2 years [6–20]. For example, infants correctly predict where agents with a false belief will look for a hidden object [7–16], and successfully take into account agents' beliefs when helping [17,18] and communicating with them [19,20]. Infants' performance in these false belief tasks have led some [6,7,21,22], although not all [23–26], researchers to argue that humans have innate cognitive machinery for forming representational relations between agents and decoupled (e.g., false) states of the world.

## No Evidence that Nonhuman Primates Use Representational Relations in ToM Tasks

Over the past few decades, researchers have also tested whether primates have the capacity to form representational relations [27–32]. Although some aspects of the performance of primates in these tasks are still debated [33–39], most researchers agree on one thing: there is currently no evidence that any nonhuman primate forms representational relations in the same way as humans. More specifically, there is no evidence that primates attribute decoupled representational states, such as false beliefs, to other agents [27–32]. In one experiment [29], chimpanzees watched as a competitor either did or did not see a high-quality food being hidden. Subjects then chose between that hidden food and a safe low-quality food. Crucially, subject chimpanzees made their own choice only after the competitor made a choice. Chimpanzees went to the hidden location for the high-quality food when the competitor had not seen that food being