

3-year-olds fail to accurately predict where a mistaken agent is likely to look for her toy if they are explicitly asked to do so. However, preverbal infants (who are not asked anything in implicit tasks) have been widely shown to expect a mistaken agent to act in accordance with the content of her false belief (cf. Baillargeon et al., 2010 for review). This is the puzzle of the discrepant developmental findings.

One of the most influential attempts at resolving this puzzle is Josef Perner's account based on the distinction between implicit and explicit understanding of another's mind. On Perner's account, young children who are unable to correctly predict the likely action of a mistaken agent when explicitly requested to do so by an experimenter lack an explicit understanding of others' false beliefs. Infants have an implicit understanding, but no explicit understanding, of others' false beliefs. What experiments based on infants' looking behavior show, on Perner's view, is that they can implicitly represent the content of another's false belief and on this basis form expectations about her likely behavior. But, unlike beliefs about others' false beliefs, infants' mere representations cannot guide their own social intentional actions: even if they spoke, they could not correctly predict the likely action of a mistaken agent when explicitly requested to do so. Nor could they intentionally help a mistaken agent (cf. Clements and Perner, 1994, 2001).

One serious challenge for Perner's account is a famous study by Buttelmann et al. (2009) who provide evidence that 18-month-olds used their representation of the content of an agent's false belief in order to actively *help* the agent find her toy (cf. Carruthers, 2013). [\[1\]](#) Thus, it is understandable that Priewasser, Rafetseder, Gargitter and Perner (2017) have recently challenged Buttelmann and colleagues' (2009) findings.

In Buttelmann et al.'s (2009) original study, an agent is facing a pair of opaque boxes A and B (a pink and a yellow box), one of which contains a toy, and the other which is empty. In the false-belief (FB) condition (where the agent falsely believed that the toy was still in the box where she had last placed it before it was moved in her absence), Buttelmann and colleagues found that toddlers reliably helped the agent find the toy by opening the non-empty box (which the agent was not trying to open). In the true-belief (TB) condition (in which the agent had been present when the toy was moved from one box to the other), the toddlers reliably helped the agent open the empty box (which the agent was trying to open). Buttelmann and colleagues take the different helping behaviors displayed by young children in the FB and TB conditions as evidence that 18-month-olds can represent the content of the agent's true or false belief.

Priewasser and colleagues' challenge is two-tiered. On the one hand, they offer an alternative interpretation of Buttelmann and colleagues' findings, based on a teleological rather than a mentalistic understanding of others' actions. According to Perner and Roessler, 2010, 2012 and Roessler and Perner, 2013, teleology makes an agent's action intelligible in terms of objective reasons, i.e. non-represented facts, not in terms of subjective reasons, which are mental representations of facts. Moreover, Priewasser and colleagues report new findings, which they take to refute the prediction based on Buttelmann and colleagues' mentalistic interpretation of their own previous findings.

1. A non-mentalistic alternative

1.1. The behavioral asymmetry between the FB and the TB conditions

If correct, Priewasser et al.'s (2017) non-mentalistic account of toddlers' helping behavior would seem to be more parsimonious than the mentalistic account in that it does not require toddlers to

represent the content of the agent's relevant (true or false) belief. This non-mentalistic account rests on three significant differences between the FB and the TB scenarios highlighted by Priewasser et al. (2017). In the FB scenario, but not in the TB scenario, the experimenter "sneakily" moves the toy from one box to the other in the agent's absence, which suggests that (i) the experimenter is "playing a trick" on the agent; (ii) the agent owns the toy, and (iii) the agent will be strongly motivated to find her toy. According to the non-mentalistic account, in the FB condition, when the agent tries to open the empty box, it is clear to young children that her goal is to find her toy. As a result, they help the agent fulfill her goal by opening the non-empty box. In the TB condition, it is less clear than in the FB condition that the agent's goal is to find the toy. When the agent unsuccessfully tries to open the empty box, her goal might instead be to open it for some unknown reason. As a result, they help the agent fulfill her goal by opening the empty box. Priewasser and colleagues argue that their account "provides teleological reasons for children to show a distinct helping pattern in the two conditions that are not based on belief reasoning." In a nutshell, their non-mentalistic claim is that they can account for the children's distinct pattern of helping without positing that young children must represent the content of the agent's *belief* in each condition.

1.2. How non-mentalistic is the non-mentalistic account?

Priewasser et al.'s claim raises two separable issues, the first of which is that in the FB scenario children are required to understand that the experimenter is "playing a trick" on the agent. The question is whether children could understand that the experimenter is playing a trick on the agent without understanding that the experimenter's goal is to cause the agent to have a false belief. As Priewasser et al. further put it (p. 3), "when [the agent] is looking for the toy in the *wrong* box children have good reason to help her find the toy in [the other] box... When she is trying to open box A *children recognize her error and correct* her by redirecting her to her toy in box B" [my emphasis]. How could children represent the empty box as the *wrong* box, let alone recognize the agent's error and correct it, unless they could represent the content of the agent's false belief?

1.3. How teleological is the non-mentalistic account?

Secondly, the question is: to what extent could the non-mentalistic account of the toddlers' helping behavior rest on the toddlers' teleological understanding of the agent's action? Teleology in the sense of Perner and Roessler (2010, 2012) and Roessler and Perner (2013) is a mode of understanding of others' intentional actions primarily based on others' objective reasons for their actions, at the expense of their subjective reasons. The hypothesis that most children before 4,5 years of age are teleologists has been put forward to explain why they fail explicit FB tasks about object-location. They fail when explicitly asked to predict the mistaken agent's likely action for two related reasons: the question itself makes sense only if the agent's action is intentional; in order to answer it children must appreciate the location where it would be rational for the agent to look for her toy. However, young children primarily think of an agent's intentional action in terms of objective reasons provided by "worldly" facts and find it difficult to distinguish the agent's objective reasons from her subjective reasons (based on the contents of her mental states, in particular her beliefs). Could young children's ability to attend to the agent's objective reasons at the expense of her subjective reasons shed light on the children's helping behavior? There are three related reasons for skepticism.

(i) It is unclear in what sense the three-point contrast between the FB and the TB scenario — in particular the fact that the agent can naturally be represented as the owner of the toy in the FB scenario, but not in the TB scenario — fits the teleological understanding of others' actions. (ii) What is distinctive of the teleological understanding of another's action is that it is based on a representation of the agent's objective reasons rather than on the agent's subjective reasons. In the TB scenario, it is a puzzle what is the agent's objective reason for opening the box she knows to be

empty. Maybe there is one, but it is unclear which one it is. In the FB scenario, only the agent's subjective reason (her false belief) for opening the empty box is manifest, not her objective reason. In fact, if young children are teleologists, they should be mystified by the agent's action in the FB condition precisely because it lacks objective reasons. If so, then it is quite unclear how they could provide efficient help to the mistaken agent. (iii) In the FB condition, the toy's actual location, which is known by the children, far from being an objective reason for the agent's action, affords an objective reason for the children's *own* action of helping the mistaken agent. But teleology is supposed to shed light on young children's understanding of an agent's objective reasons, at the expense of her subjective reasons, *not* on the children's own objective reasons for their *own* actions. All of this, I think, casts serious doubt on the claim that the "non-mentalistic" alternative is a genuine instance of a teleological understanding of the agent's action.

2. The refutation of the mentalistic interpretation

Priewasser and colleagues have not merely offered a tentative non-mentalistic alternative to the mentalistic interpretation of Buttelmann et al.'s (2009) findings. They have also offered new evidence that they claim refutes the prediction derived from the mentalistic interpretation.

They have run a new pair of FB and TB conditions, involving not only two boxes (as in the previous study), but three boxes A, B and C. In both the new FB and the new TB conditions, after the agent's toy has been moved by the experimenter to box B (either in the absence or the presence of the agent), the agent now unsuccessfully tries to open the third box C, not box A, in which the agent first placed her toy.

Priewasser and colleagues report that in the new FB condition most children reliably helped the mistaken agent by opening the non-empty box B that contained the toy and in the new TB condition, most of them reliably helped the agent open the empty box C that she was unsuccessfully trying to open. They plausibly argue that since the early stages of the new FB scenario are similar to the early stages of the old FB condition, children in the new FB condition, just like in the old FB condition, should be expected to open the non-empty box B in order to help the mistaken agent find her toy. They also argue that in the new TB condition, just like in the old TB condition, children should be expected to help the agent open the empty box that she is unsuccessfully trying to open, namely box C. They further argue that the mentalistic interpretation should make the same prediction about children's helping behavior in the new TB condition, but not in the new FB condition. As they put it (p. 3), "for the new-FB condition the two theories make different predictions. If children use [the agent's] belief and knowledge to infer what she wants..., then children should behave in the new-FB condition in the same way as in the original TB condition: they should help open box C, since [the agent] knows that *box C does not contain her toy anymore* and she cannot be looking for it" [my emphasis]. Since the findings show that in the new FB condition, most children reliably helped the mistaken agent by opening box B, not box C, they take their findings to refute the mentalistic prediction.

When Priewasser and colleagues try to justify their claim that the mentalistic prediction is committed to treating the new FB condition on a par with the old TB condition, they characteristically write that children "should help open box C, since the agent knows that *box C does not contain her toy any more*." Is this a mistake (a 'thinko') or a typo? Box C never contained the agent's toy at all. This is a major difference between the old TB and the new FB condition: in the TB condition, the agent knows that box A does not contain her toy any more. But in the latter, the agent cannot be said to know that box C does not contain her toy anymore, since it never did. Whether it is a mistake or a typo, it casts doubt, I think, on Priewasser and colleagues' grounds for imputing to the mentalistic interpretation the prediction that children should behave in the new FB condition in

the same way they behave in the old TB condition.

But I see no convincing reason why the mentalistic prediction should accept the burden of assuming that children in the new FB condition should behave as they did in the old TB condition. Nor do I see any reason why the mentalistic prediction should be prohibited from making good use of the three-point contrast between the TB and the FB condition highlighted by Priewasser and colleagues. There are at least three differences between the old TB condition and the new FB condition, the first of which is the main bone of contention between advocates and critics of the mentalistic account, namely the difference between the agent's having a true and a false belief.

The second difference is that in the old TB condition, the agent first placed her toy in box A, not in box B, before it was moved in her presence by the experimenter to box B. The fact that the agent selected box A to place her toy in the old TB condition is consistent with her having some unknown motivation to open box A. When the agent unsuccessfully tries to open box A while knowing that her toy is in box B, young children may assume that she has some reason or other for trying to open box A based on her prior selection of box A over box B for placing her toy, although they do not know her reason. By contrast, in the new FB condition, when the mistaken agent tries unsuccessfully to open box C, children cannot draw on the fact that the agent earlier placed her toy in box A (not C), in order to infer that the agent must have some unknown reason to deal with box C (in which she never earlier placed her toy).

Finally, as Priewasser and colleagues rightly emphasize, there is a third relevant difference between the early stages of respectively the old TB condition and the new FB condition. Only the early stages of the FB conditions (not the early stages of the TB conditions) are consistent with the assumption that the agent owns the toy and is greatly motivated to find it.

In light of the second difference between the old TB and the new FB condition, the mentalistic account is likely to predict that the children will be baffled by the fact that the agent's attempt at opening box C cannot be justified by her false belief that her toy is in box A (as the agent's action was in the old FB condition). They will also be more baffled by the agent's action in the new FB condition than by the agent's action in the old TB condition. In light of the fact that in the new FB condition (but not in the old TB condition), the agent holds a false belief about her toy's location and is also naturally construed as eager to find the toy that she owns, the children are likely to reason that if their goal is to help the agent, then the most efficient means at their disposal is to provide her with her toy (about whose location she has a false belief).

In this short note, I have argued that the non-mentalistic account of the findings based on active helping (put forward by Priewasser and colleagues) does not comfortably count as an instance of teleology (in the sense of Roessler and Perner, 2013). I have further argued that their new findings do not squarely refute the mentalistic prediction. [\[2\]](#)

References

- Baillargeon, R., Scott, R. M., & He, Z. (2010) False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009) Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2), 337-342.
- Carruthers, P. (2013) Mindreading in infancy. *Mind & Language*, 28(2), 141-172.

Clements, W. A., & Perner, J. (1994) Implicit understanding of belief. *Cognitive Development*, 9, 377-397.

Clements, W. A., & Perner, J. (2001) When actions really do speak louder than words-but only implicitly: young children's understanding of false belief in action. *British Journal of Developmental Psychology*, 19, 413-432.

Perner, J., & Roessler, J. (2010) Teleology and causal reasoning in children's theory of mind. In J. Aguilar, & A. Buckareff (Eds.). *Causing human action: New perspectives on the causal theory of action* (pp. 199-228). Cambridge, MA: Bradford Book, The MIT Press.

Priewasser, B., Rafetseder, E., Gargitter, C. & Perner, J. (2017) Helping as an early indicator of a theory of mind: Mentalism or teleology? *Cognitive Development*, <http://dx.doi.org/10.1016/j.cogdev.2017.08.002>

[1] In Butteltmann et al.'s (2009) study, the agent was male. However, I will refer to a female agent in accordance with the study by Priewasser et al. (2017), in which they first replicated the study by Butteltmann and colleagues before providing new data.

[2] Thanks to Gyuri Gergely and Dan Sperber for their comments.