

Commentary Accepted at PNAS 18th January 2018.

Conceptualizing Degrees of Theory of Mind

Jane Rebecca Conway^{a,b} and Geoffrey Bird^{a,b,1}

^a Department of Experimental Psychology, University of Oxford, New Radcliffe House, Radcliffe Observatory Quarter, Oxford, OX2 6GG, United Kingdom.

^b MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, Denmark Hill, London, SE5 8AF, United Kingdom.

¹geoff.bird@psy.ox.ac.uk

Successful navigation of the social world requires making accurate inferences about the contents of other people's minds, being able to represent in one's own mind the thoughts, beliefs, and intentions of another. This "theory of mind" (ToM) ability allows us to explain and predict others' behavior in terms of their mental states (1). As reported in PNAS, Bio et al. (2) show that when prompted to adopt the visual perspective of a cartoon agent participants demonstrated the same spatial bias as when processing objects from their own perspective. Interestingly, this effect occurred only when the cartoon agent held a false belief about the location of an object, due to its having moved while the agent's view was blocked by a barrier. When the agent held a true belief about the location of the object, either because he had witnessed it move or it never moved from its original location after the barrier appeared, participants' own spatial bias had no effect on perspective taking (see Fig. 1A for a description of the experimental paradigm). Bio et al. conclude that when representing others' true beliefs, compared with their false beliefs, social cognition may be engaged to a lesser extent or not at all. This study is important because it may provide an answer to one of the most challenging questions in current ToM research: What is it to represent mental states to a greater or lesser degree?

Despite 40 years of ToM research (1), two related questions remain unanswered. The first concerns individual differences in ToM ability: What does it mean for an individual to be "better" at ToM than another individual? There is substantial evidence that individual differences in general cognitive abilities such as intelligence, language, or memory (3–7), have an impact on ToM tasks, but such skills are not specific to mental-state representation and would improve performance on any task. The second relates to ToM tasks: What does it mean for one task to

require a “greater degree” of ToM than another? Perhaps the most successful effort to date in determining a metric by which to judge the difficulty of ToM processing has been a scale developed by Wellman and Liu (8) that is derived from the order in which children typically are able to represent different categories of mental states. For instance, children tend to understand that other people can hold different beliefs from their own before being able to ascribe false beliefs to them. However, the order of successfully representing different mental-state categories differs across cultures (9, 10), indicating that variability may depend more on what children are taught about minds (11) rather than providing any objective measure, or explanation, of the degree to which a particular type of ToM representation is more difficult than another.

These problems intersect when trying to interpret performance on tests of ToM ability; for each type of mental-state representation, ToM is presented as a pass-or-fail ability – children either can represent mental state type X, Y, or Z, or they cannot. Such binary responses cannot capture variance in ToM, or what it is to engage in mental-state representation to a greater or lesser degree. After children acquire the ToM ability for each category, typically by the age of five (12), it is not clear what “more” or “better” ToM is. There is a lack of theorising on variance in ToM (between tasks and between individuals) that can address the differential quality of mental-state representations. Within this context it is extremely interesting to consider the novel task developed by Bio et al. (2).

Bio et al. employ a task designed to evoke two mental-state categories, true vs. false beliefs. In contrast to false beliefs, true beliefs have been considered problematic as a ToM measure (13, 14) as their representational content is the same as the true

state of the world, making it impossible to tell whether one is truly representing another person's mental state or simply reporting one's own representation of reality. Bio et al. suggest that their findings indicate that ToM was engaged and required only, or to a greater degree, in the false-belief condition. They suggest that on true-belief trials participants may have shown no spatial bias as they did not represent the agent's mental state, or did so to a lesser degree, and therefore did not represent the agent's visual perspective. Based on this explanation, individual differences in ToM may be construed as the extent to which representing the cartoon as a mental agent prompted one to adopt his visual perspective. If so, this represents an interesting new measure of individual differences in ToM. In contrast to existing measures, the dependant variable does not reflect whether one can represent particular classes of mental states, or the accuracy of mental state inference, but rather the social consequences (visual perspective taking) of false-belief attribution. When construed in this manner, it is interesting to speculate whether those with a "higher" degree of ToM, or a greater propensity to represent the mental states of others, may also show such an effect on true-belief trials even when the representation of the agent's mental state is not strictly necessary for task performance. If so, one might expect a correlation between the propensity to engage in ToM and the difference in spatial bias between true- and false-belief conditions.

On false-belief trials, participants were prompted to respond from another agent's visual perspective when it conflicted with their own. This was demonstrated by the fact that the participant's own degree of horizontal left-right spatial bias was evident in their responses to what were vertically arranged target locations from their perspective, but horizontally arranged target locations if they adopted the agent's

perspective. It is interesting that the authors suggest that participants may have attributed their own spatial bias to the agent, as egocentric projection in ToM often indicates a less accurate representation of another's mind (15). In this task, as in many ToM tasks, one is asked to infer the mental state of an anonymous agent in the absence of any knowledge about his mind. When required to represent the mental states of such an agent, there are at least two strategies one could take: One could either predict the mental state of the average mind in that situation or predict one's own mental state in that situation. When these conflict, probabilistically one is more likely to be accurate if one represents the mental state of the average mind. Accordingly, perhaps those participants who are most accurate at representing the spatial bias of the cartoon agent are those who do not attribute their own spatial bias to the agent, but instead represent the veridical average degree of spatial bias in this sample (no spatial bias), or that commonly observed in the previous literature (a left spatial bias (2)).

One further possibility is that instead of representing the cartoon agent's visual perspective, participants were instead representing the spatial arrangement of the stimuli from a position centred on the agent's viewpoint. The former case would have mentalistic representational content and be mediated by social cognitive processes. In contrast, the latter case would not require any representation of the agent's mental state and be mediated by nonsocial object-centred spatial processing. As an example, consider the case of a guest at a wedding who wants to ensure she is visible in the wedding photograph. She could do so either by imagining the perspective of the photographer through the camera lens (a mentalistic process), or by calculating whether there is an unobstructed straight line between her and the camera lens (a

nonmentalistic process). Whether performance is driven by representations of mental states on specific tasks is a long-standing debate in the ToM literature, particularly for ToM tasks where the responses are nonverbal and relatively fast (16, 17).

Dissociating mentalistic from nonmentalistic accounts of nonverbal ToM tasks is typically achieved by including a nonsocial control condition, in which the agent is replaced by an inanimate object to which it is not appropriate to attribute mental states (18, 19). Cameras are frequently used in tasks that manipulate beliefs on the basis of visual representations, as photographs provide visual representations of a scene but are not mental states, and can, like beliefs, become “false” representations of reality with the passage of time (20). See Fig. 1B for an example of how such a nonsocial camera control condition could be incorporated into the paradigm used by Bio et al. in future research. Such further work could establish whether, as currently hypothesised, participants may not have fully represented the agent’s true beliefs, or whether in fact no belief representation occurred on either true- or false-belief trials.

While any new task prompts a variety of questions about the factors that determine performance, Bio et al. (2) are to be congratulated on conceiving of an entirely novel approach to understanding individual differences in ToM. Rather than the accuracy of mental-state inference, the ability to represent different types of mental state, or the propensity with which one engages in mental-state reasoning, their novel task attempts to measure the consequences of mental-state representation on other social cognitive processes. Whether this approach proves fruitful will only be determined by further empirical and theoretical work, but their paper provides a possible glimpse of the solution to the deep problem of how to conceive of degrees of ToM, a problem which has vexed the field since its inception.

Acknowledgements

GB and JRC gratefully acknowledge the support of the Baily Thomas Trust, and the Economic and Social Research Council.

References

1. Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 4:515–526.
2. Bio B, Webb T, Graziano MSA (In Press) Projecting One' s Own Spatial Bias Onto Others During a Theory-of-Mind Task. *Proceedings of the National Academy of Sciences*.
3. Ronald A, Viding E, Happé F, Plomin R (2006) Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: a twin study. *Social Neuroscience* 1(3–4):412–25.
4. Milligan K, Astington JW, Dack LA (2014) Language and Theory of Mind : Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development* 78(2):622–646.
5. Brüne M (2003) Theory of mind and the role of IQ in chronic disorganized schizophrenia. *Schizophrenia Research* 60:57–64.
6. Hughes C (1998) Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology* 16(2):233–253.
7. Hughes C, et al. (2005) Origins of individual differences in theory of mind: from nature to nurture? *Child Development* 76(2):356–70.
8. Wellman HM, Liu D (2004) Scaling of Theory-of-Mind Tasks. *Child Development* 75(2):523–541.

9. Shahaecian A, Peterson CC, Slaughter V, Wellman HM (2011) Culture and the Sequence of Steps in Theory of Mind Development. *Developmental Psychology* 47(5):1239–1247.
10. Slaughter V, Perez-Zapata D (2014) Cultural Variations in the Development of Mind Reading. *Child Development Perspectives* 8(4):237–241.
11. Heyes CM, Frith CD (2014) The cultural evolution of mind reading. *Science* 344(6190):1243091.
12. Wellman HM, Cross D, Watson J (2001) Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development* 72(3):655–684.
13. Dennett DC (1978) Beliefs about beliefs. *Behavioral and Brain Sciences* 4:568–570.
14. Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a “theory of mind”? *Cognition* 21(1):37–46.
15. Apperly IA, et al. (2010) Why are there limits on theory of mind use? Evidence from adults’ ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology* 63(6):1201–1217.
16. Heyes C (2014) Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science* 9(2):131–143.
17. Heyes C (2014) False belief in infancy: a fresh look. *Developmental Science* 17(5):647–659.
18. Santiesteban I, Shah P, White S, Bird G, Heyes C (2014) Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review* 22(3):844–849.
19. Conway JR, Lee D, Ojaghi M, Catmur C, Bird G (2017) Submentalizing or

mentalizing in a level 1 perspective-taking task: A cloak and goggles test.

Journal of Experimental Psychology: Human Perception and Performance

43(3):454–465.

20. Saxe R, Kanwisher N (2003) People thinking about people: The role of the temporo-parietal junction in “theory of mind.” *NeuroImage*, 19(4):1835–1842.

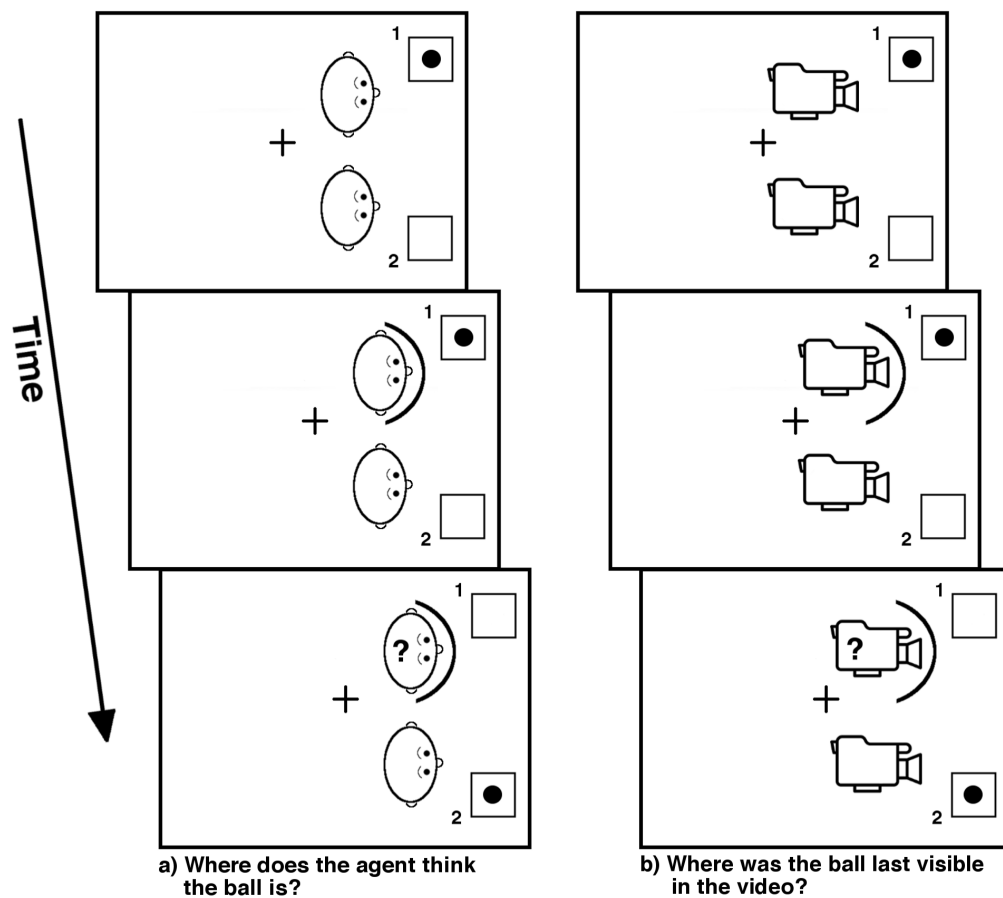


Fig. 1. A social “false-belief” condition (A) in the task by Bio et al. (2) and an analogous nonsocial “false-photo” condition (B). Two agents view (A) and two video cameras record (B) a scene in which a ball appears in one of two boxes. An occluder then appears between one of the agents/cameras and the ball, before it moves to the other box. Participants are asked to indicate in which box the agent thinks the ball is (social version, A) or where the ball was last visible in the video (nonsocial version, B). In their social version, Bio et al. (2) found participants responded faster to either box 1 or 2, situated on the left and right of the agent’s head, depending on their own bias in processing either the left or right side of space. Comparing social and nonsocial conditions could distinguish between mentalistic and nonmentalistic explanations of this effect.