Journal: Psychological Assessment

Format: Commentary article

Word count: 2821

Date: August 2024

Theoretical Limitations on Mindreading Measures: Commentary on Wendt et al. (2024)

Jane R. Conway*[1], Emily L. Long[2], Leora Sevi[2], Caroline Catmur[3] & Geoffrey Bird[2,4]

1. School of Psychology, University of Galway, Galway, Ireland.
2. Department of Experimental Psychology, University of Oxford, Oxford, UK.
3. Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.
4. Centre for Research in Autism and Education, Institute of Education, University College London, London, UK.

Corresponding author: Jane R. Conway, jane.conway@universityofgalway.ie

Theoretical Limitations on Mindreading Measures: Commentary on Wendt et al. (2024)

Abstract

In this *Commentary* article we expand on issues in the Theory of Mind literature raised by Wendt et al. (2024) that limit progress in our understanding of how people read other minds. We critically assess how they categorized tasks in their study, and in so doing raise deeper questions that need addressing: what exactly are *mental states;* how can we accurately measure mindreading when the ‘correct’ answer lacks ground truth; and what are the contributions to individual differences in mindreading of general cognitive ability and specific experience in the kinds of minds being read? We conclude that developing a psychological theory of *how* people read other minds would advance ways in which we can better measure and explain what it means to be better or worse at mindreading, and how general cognitive ability relates to this socio-cognitive skill.

Keywords: theory of mind, Mindspace, social cognition, individual differences, measurement

Public Significance Statement: Researchers do not yet have a good understanding of why people differ in their ability to understand what others are thinking (‘theory of mind’). We highlight how developing a psychological theory of how people understand others’ minds will allow this understanding to be measured, and the relationship between theory of mind and other more general types of intellect to be understood.

Theoretical Limitations on Mindreading Measures: Commentary on Wendt et al. (2024)

Wendt et al. (2024) address an important issue in the mindreading (or Theory of Mind; ToM) literature by examining the relationships between individuals' performance on tasks purporting to measure ToM, their performance on self-report measures of ToM ability, and their general cognitive ability and psychological functioning. Their article is scholarly, thoughtful, and is admirable in contrasting rival hypotheses with empirical data. While valuing immensely the data and thinking in Wendt and colleagues' paper, we would like to expand upon some of the issues they raised for the study of theory of mind or mindreading, as well as address an inaccuracy in how they describe a test we developed. In our view one of the biggest problems in the study of ToM is the lack of a psychological theory as to *how* humans 'read minds'. The other big problems are how researchers measure ToM ability; and defining what mental states actually are. The paper by Wendt and colleagues highlights how some of these issues interact to cause problems for the field and are the basis of our commentary.

The lack of a psychological, mechanistic, theory of how we make mental state inferences (or how we 'read minds') is problematic for a number of reasons. Most pertinent to the current commentary is that, without such a theory, it is hard to determine what one would expect the relationship between ToM and general cognitive ability to be under what Wendt and colleagues call the validity hypothesis – that both "*self-report and task-based measures of mindreading ability are valid… that their test scores capture mindreading ability without simultaneously reflecting unrelated traits or other contaminating influences*". Without such a theory it is hard to know whether other traits (or abilities) are related to mindreading, unrelated to mindreading, constituent processes of mindreading, or a contaminating influence.

We will briefly outline here our attempt to develop a mechanistic model of how we make mental state inferences (for a full account see Conway et al., 2019), as it directly relates to the issues at hand. Under the Mindspace framework, we suggest that in order to make mental state inferences (i.e., to mind read, or to work out what another is thinking) individuals calculate the likelihood of various mental states given (at least) two things – the situation the target individual is in, and the characteristics of their mind. For example, it is fairly intuitive that the content of mental states a target individual will have at a party will depend on whether they are an introvert or an extrovert, and indeed our published work suggests that characteristics of mind are taken into account by individuals when inferring mental states (Conway et al., 2020; Long et al., 2022). The theory predicts that there are a variety of factors that should explain individual differences in theory of mind ability. These include (but are not limited to) one's ability: to understand how minds vary within the population, to determine the characteristics of a target's mind, to understand how situations predict the likelihood of different mental states, and to combine these sources of information to make inferences about the contents of mental states given a specific mind in a specific situation.

The importance of the first factor – the ability to understand how minds vary – arises from the prediction that certain features of minds, such as personality traits, will have different degrees of diagnosticity with respect to the likelihood of specific mental states in specific situations, and information about these features may be more or less available in the social stimuli. For example, if extraversion is the most diagnostic trait when required to determine a target individual's mental state on a particular occasion, but I only have information on the target's level of conscientiousness, then I will make a better mental state inference if I know the relationship between conscientiousness and extraversion, and so can make a prediction about the target's level of extraversion from their level of conscientiousness. This knowledge is what is tested by the Personality Pairs Task, included by Wendt and colleagues in their study as a test of ToM. The Personality Pairs Task, as correctly described by Wendt and colleagues, involves participants estimating the likelihood that pairs of person-descriptive statements from the HEXACO-100 (Lee & Ashton, 2018) are both true for the same person. Importantly, this task does *not* measure ToM, defined as the accuracy of mental state inferences. It measures the understanding of how minds (co-)vary along six personality dimensions. However, individual differences on this task should predict the accuracy of mental state inferences in general, and so there should be high correlations between this task and tasks that measure ToM (where ToM ability is defined as the accuracy of mental state inferences). The simplest explanation for why this should be (ignoring other factors that might cause correlations between scores on the Personality Pairs Task and tests of ToM ability, such as general skill in social learning, degree of social attention, social experience, etc.) is that under the Mindspace framework, the process of trait inference is (often) a constituent part of making mental state inferences. High accuracy in the Personality Pairs Task, then, is thought to be reflective of an understanding of population trait covariation that facilitates accurate trait inference, which itself is useful in making accurate mental state inferences.

Under the Mindspace framework, mental state inferences are no different from any other inferences which rely on the combination of multiple sources of information (e.g., the situation a target individual is in and the characteristics of their mind). Thus, to the degree that general cognitive ability or IQ determines the speed and/or accuracy of inferences in general, so it will with mental state inferences. However, as often occurs in non-mentalistic reasoning tasks, the accuracy of inferences is determined by specialised knowledge – about how minds vary, how observed behaviours map onto psychological traits, and the likelihood of specific mental states in specific social situations. This is analogous to a non-mentalistic reasoning task such as predicting the likelihood of a storm at sea, where the accuracy of inferences depends on specialist knowledge of local weather patterns and how barometric pressure predicts weather. Thus, the accuracy of mental state inferences will be determined by cognitive ability to a degree, but also by the accuracy of the knowledge used to inform such inferences. However, the accuracy (or extent) of supporting knowledge used to make inferences may itself be determined in part by general cognitive ability, as this knowledge must be learned from experience – meaning that one might expect high correlations between general cognitive ability and mental state inference accuracy.

If these correlations do exist, Wendt and colleagues reasonably ask whether we should consider ToM an epiphenomenon of general cognitive ability, or whether general cognitive ability tasks might be sufficient to assess ToM ability. In our view the answer is no to both questions, because, at least under the Mindspace framework, they may dissociate in specific individuals despite showing a degree of association across individuals. For example, an individual may have a very high IQ but lack the relevant social knowledge, or knowledge about other minds, to make accurate mental state inferences. In fact, under the Mindspace framework, it is very possible that an individual's ToM ability (more formally the accuracy of their mental state inferences) varies according to the target mind and the situation that Target is in. For example, an individual who has grown up in the UK surrounded by a relatively homogeneous group of individuals may be adept at inferring the mental state of others like her in situations she is accustomed to, but poor when inferring the mental states of a Japanese target mind in an unfamiliar situation. From this example two things can be seen: 1) a dissociation between relative ToM ability and general cognitive ability may *explain* individual differences in ToM ability; and, 2) it may be advantageous to think of a general ToM ability (influenced by social learning ability; social attention; and quantity, quality and diversity of social experience) which may be determined in part by general cognitive ability, as existing alongside specific ToM abilities to 'mind read' certain target individuals, or groups of targets, which are more governed by experience.

Task-based tests of ToM, or mindreading ability, have two problematic characteristics in our view, which limit the conclusions one can make from the excellent work conducted by Wendt and colleagues. The first is that they often ask participants to infer the mental states of anonymous protagonists, about whom nothing is known. If real life skill in inferring the mental states of others relies on being able to judge the characteristics of their mind, and knowing how these characteristics interact with the situation a target individual is in to determine their mental states, then it is unsurprising that individuals' self-reports about their real life ToM ability do not correlate well with performance on tests which do not include these crucial components of ToM ability. Moreover, whilst such tests, by excluding these aspects of ToM ability, may be thought to provide a measure of one's understanding of the mental states held by the 'average' mind, they face the additional issue of whether the supposed average on which the task is based aligns with the average individual from the participant's social environment. Second, a basic requirement of a test of ability is that accuracy can be judged – that we know whether a participant got a question right or wrong. Problematically, until the Interview Task was developed (Long et al., 2022), as far as we are aware no task was able to fulfil this basic requirement (although see Moritz & Roberts, 2020 for a interpersonal liking task that compared self and other judgements with ground truth data). Consider the Movie for the Assessment of Social Cognition (Dziobek et al., 2006; one of the measures included by Wendt and colleagues). In this task participants watch a scripted, fictional drama, and are asked about the characters' mental states. Here the 'correct' answer is that determined appropriate by the script writer – the characters themselves have no actual mental states as they aren't real, only the actors playing the characters do, and we do not know what the actors' mental states were. The same is true for the Reading the Mind in The Eyes Test (Olderbak et al., 2015; the short form of which was also included by Wendt and colleagues): the correct answer is determined by consensus, we do not know the actual mental states of the individuals depicted in the stimulus materials. Since we do not know what the correct answer is in these tests, and indeed it might not be possible for there to be a correct answer in tests using fictional characters, it is unsurprising that self-report measures asking respondents to report their level of ToM ability in real life do not correlate with tests which are unable to test their ability.

Using the Interview Task (Long et al., 2022; a task in which participants watch real-life social interactions and infer the mental states of interactants where 'ground truth' about those mental states is known) one *can* separate out a general ToM ability from target-specific abilities, and assess the relationship between these sources of variance in performance and general cognitive ability, as well as measure self-reported perception of performance via confidence judgements (Long et al., *under revision*). Given this possibility, it is useful to consider issues raised by Wendt and colleagues' results. If one considers a standard ToM task in which participants read or watch vignettes about fictional characters and are asked to infer their mental states, it is relatively unsurprising to find a correlation between this task – indeed between almost any task – and general cognitive ability if one avoids floor and ceiling effects and task performance is not too dependent on personality factors such as conscientiousness etc. The classic approach to 'solve' this problem (if it needs solving) would be to contrast the performance on the mental state vignette task with performance on a control task in which participants read vignettes and make non-mentalistic inferences, with the idea that if both tasks are matched for general cognitive demands, then the performance component unique to mental state inferences would be distinct from general cognitive ability. The tasks used by Wendt and colleagues either do not have such control tasks, or the control tasks are not well-matched to the mentalistic versions. As such, even ignoring the previously-described issues with the tests, if one assumes that ToM ability is the result of the performance of a distinct and relatively isolated 'module', these tests cannot measure that performance independently of general cognitive ability.

As mentioned previously, however, under the Mindspace framework it is not clear what controlling for general cognitive ability would achieve. The framework allows that general cognitive ability is likely to explain a large amount of variance in the speed and accuracy of mental state inferences, and might explain a large amount of variance in the accuracy and extent of social knowledge about situations and minds used to make those inferences. However, the degree to which it does so will depend on experience of the social situations and target minds which are being assessed. Thus, the variance unaccounted for by IQ will likely vary according to test and target population, their interaction, and individual-level experience. This can be intuited if one considers the relationship between general cognitive ability and academic performance in schools – performance will depend on personality traits of the student, the ability of their teachers, the length of time they have been studying a particular subject, and yet general cognitive ability will still contribute to performance within wide ranges of all these factors to a greater or lesser extent (e.g., a high IQ will not allow you to perform well on a test of the Icelandic language if you have never studied it).

An additional complication relates to the question of *what is a mental state*? Across tasks mental states are considered to be thoughts (Dziobek et al., 2006), propositional attitudes (Hughes et al., 2000), emotions (Oakley et al., 2016), visual perspectives (Santiesteban et al., 2015), or even states of mind (Tamir & Thornton, 2018). Since skill in inferring these different types of mental state may dissociate in task-based measures (Conway et al., 2017; Oakley et al., 2016; Santiesteban et al., 2015; and note that if emotions are considered to be mental states then so-called empathic accuracy tasks, (Lantos et al., 2023; Santiesteban et al., 2021), are valid measures of ToM ability with a ground truth accuracy rather than measures of empathy (Coll et al., 2017), it could be argued that self-report measures of ToM need to be designed such that they assess ability in inferring each of these types of content independently. Furthermore, if it is accepted that ToM ability might depend on the specific target mind and the situation it is in, then self-report questionnaires should be matched to the targets and situations assessed by the task (an alternative valid approach would be to measure confidence in task performance as also used by Wendt and colleagues).

Finally, given that appropriate task-based and self-report measures will never measure the same thing (they measure ability versus perception of ability, respectively), one would only expect these measures to correlate in those with an appropriate degree of self-awareness or high metacognitive ability.

Finally, the excellent paper by Wendt and colleagues raises interesting questions about what it means for a test to be valid, putting forward the possibility "that the pursuit of perfectly valid measures of mindreading ability may be an unrealistic goal." It is possible that under some conceptualisations, validity may be being confused with being 'process pure'. Unless one believes that ToM is performed by a 'Theory of Mind Module' which accomplishes *all* of the psychological work necessary to make mental state inferences, then the accuracy of mental state inferences will be determined by a number of psychological processes and different sources of knowledge. Controlling for one or a number of these processes / sources of knowledge will change the interpretation of test scores, and may result in any individual's performance relative to others changing dramatically. In our view, however, this would not mean that tests of ToM ability are invalid. If a test requires participants to infer the mental states of individuals, and those mental states are known so that the accuracy of mental state inferences can be calculated, then the test is a valid test of ToM, regardless of whether ToM is drawing on a number of psychological processes.

Overall, therefore, we strongly agree with Wendt and colleagues that the validity of mindreading or ToM tasks is an area of psychological assessment that requires urgent attention; however, it is important to ensure that the tasks being studied are indeed designed to measure mindreading. Furthermore, assessment of the validity of mindreading tasks must address the issue that most tasks do not measure the accuracy of mindreading against a true correct answer.

References

Coll, M.-P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews*, *83*, 132–139. https://doi.org/10.1016/j.neubiorev.2017.10.009

Conway, J. R., Catmur, C., & Bird, G. (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychonomic Bulletin & Review*, *26*(3), 798–812. https://doi.org/10.3758/s13423-018-1559-x

Conway, J. R., Coll, M.-P., Cuve, H. C., Koletsi, S., Bronitt, N., Catmur, C., & Bird, G. (2020). Understanding how minds vary relates to skill in inferring mental states, personality, and intelligence. *Journal of Experimental Psychology: General*, *149*(6), 1032–1047. https://doi.org/10.1037/xge0000704

Conway, J. R., 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. https://doi.org/10.1037/xhp0000319

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A Movie for the Assessment of Social Cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636. https://doi.org/10.1007/s10803-006-0107-0

Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good Test-Retest Reliability for Standard and Advanced False-Belief Tasks across a Wide Range of Abilities. *Journal of Child Psychology and Psychiatry*, *41*(4), 483–490. https://doi.org/10.1111/1469-7610.00633

Lantos, D., Costa, C., Briglia, M., Molenberghs, P., Kanske, P., & Singer, T. (2023). Introducing the English EmpaToM task: A tool to assess empathy, compassion, and theory of mind in fMRI studies. *Neuroimage: Reports*, *3*(3), 100180. https://doi.org/10.1016/j.ynirp.2023.100180

Lee, K., & Ashton, M. C. (2018). Psychometric Properties of the HEXACO-100. Assessment, 25(5), 543-556. https://doi.org/10.1177/1073191116659134

Long, E. L., Cuve, H. C., Conway, J. R., Catmur, C., & Bird, G. (2022). Novel theory of mind task demonstrates representation of minds in mental state inference. *Scientific Reports*, *12*(1), 21133. https://doi.org/10.1038/s41598-022-25490-x

Long, E. L., Catmur, C., Fleming, S.M., & Bird, G. (*Under Revision*). Metacognition facilitates Theory of Mind through optimal weighting of trait inferences.

Moritz, D., & Roberts, J. E. (2020). Depressive Symptoms and Self-Esteem as Moderators of Metaperceptions of Social Rejection Versus Acceptance: A Truth and Bias Analysis. *Clinical Psychological Science*, 8(2), 252-265. https://doi.org/10.1177/2167702619894906Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology*, *125*(6), 818–823. https://doi.org/10.1037/abn0000182

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01503

Santiesteban, I., Gibbard, C., Drucks, H., Clayton, N., Banissy, M. J., & Bird, G. (2021). Individuals with Autism Share Others' Emotions: Evidence from the Continuous Affective Rating and Empathic Responses (CARER) Task. *Journal of Autism and Developmental Disorders*, *51*(2), 391–404. https://doi.org/10.1007/s10803-020-04535-y

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review*, *22*(3), 844–849. https://doi.org/10.3758/s13423-014-0716-0

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Wendt, L. P., Zimmermann, J., Spitzer, C., & Müller, S. (2024). Mindreading measures misread? A multimethod investigation into the validity of self-report and task-based approaches. *Psychological Assessment*. https://doi.org/10.1037/pas0001310