Title

A theoretically driven meta-analysis of implicit theory of mind studies: The role offactivity

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Holland, Catherine
Phillips, Jonathan

Publication Date

2020

Peer reviewed

A theoretically driven meta-analysis of implicit theory of mind studies: The role of factivity

Catherine Holland (Catherine.A.Holland.GR@dartmouth.edu)

Jonathan Phillips (Jonathan.S.Phillips@dartmouth.edu)
Program in Cognitive Science, 23 N Main Street
Hanover, NH 03755 USA

Abstract

The capacity for Theory of Mind (ToM) allows us to represent others' understanding of the world independently from our own and then explain and predict their actions in terms of their understanding. Researchers have often focused on trying to find evidence for an *implicit* theory of mind system: one that emerges early in human ontogeny and operates mandatorially in adults. In this paper, we ask how the recent methodological push towards replication can be used as a tool that bears on a key theoretical distinction in implicit Theory of Mind, namely the distinction between factive and non-factive ToM representation. Unlike other meta-analyses, our primary interest is not the overall replicability of theory of mind findings. Instead, we ask whether the replicability of implicit theory of mind tasks depends in part on whether they measured factive or non-factive ToM. We find that, to the extent that there is replicable and robust evidence for implicit ToM, that evidence largely comes from tests that investigated factive ToM representations. This analysis is a proof of concept of the broader potential for using replication attempts to ask theoretically motivated questions.

Keywords: Theory of Mind; Factive; Non-Factive; Replication; Meta-Analysis

Theory of Mind (ToM) allows us to represent others' understanding of the world independently from our own and thus explain and predict others' actions in terms of their understanding. Researchers have argued that this capacity may emerge extremely early in human ontogeny (Southgate & Vernetti, 2014; Kovács, Téglás, & Endress, 2010) and operate automatically in human adults (Apperly & Butterfill, 2009; van der Wel, Sebanz, & Knoblich, 2014), and thus constitute an *implicit* theory of mind system.

The majority of the research seeking to provide evidence for implicit theory of mind has primarily focused on demonstrating an ability to represent others' *non-factive* mental states, e.g., beliefs in a false belief paradigm. While this emphasis on early emerging and automatic false belief representation has been highly influential, it has recently come under challenge on both empirical and theoretical grounds.

On empirical grounds, many of these results have come under scrutiny with the recent emphasis on replicating key findings in psychology. On theoretical grounds, a number of researchers have recently argued non-factive tests neglect to capture more basic aspects of theory of mind, namely representations of *factive* mental states, e.g., what others' know, see, hear, are aware of, and so on (Martin & Santos, 2016; Nagel, 2017; Phillips & Norby, 2019). Once again, the replication crisis has also scrutinized whether the original evidence for implicit *factive* theory of mind replicates.

In this paper, we ask how this recent methodological push towards replication can be used as a tool that bears on this theoretical distinction. Unlike other meta-analyses, our primary interest is not the overall replicability of implicit theory of mind tasks, or the replicability of any specific effect that has been previously demonstrated; instead, we ask whether the replicability of theory of mind tasks depends in part on whether they measured factive or non-factive theory of mind. In an ongoing meta-analysis we find that, to the extent that there is replicable and robust evidence for implicit theory of mind, that evidence largely comes from tasks that demonstrated factive theory of mind representations. This analysis is a proof of concept of the broader potential for using replication attempts to ask theoretically motivated questions.

Background

Early investigations of theory of mind

Since the skeptical reception of the seminal paper 'Does the chimpanzee have a theory of mind?' by Premack and Woodruff (1978), most researchers have emphasized the importance of using false-belief representation as a litmus test for demonstrating a theory of mind (Dennett, 1978; Bennett, 1978). To attribute a false belief, an individual must represent someone as having an understanding of the world that is strictly inconsistent with their own. This requirement ensures that correct predictions of what the other agent will do cannot be based solely on one's own representation of the world.

The most widely-used false belief paradigm is referred to as a Sally-Anne test (Baron-Cohen, Leslie, & Frith, 1985). In this paradigm, Sally puts a toy in a basket in the presence of her friend Anne. Sally then leaves the room, during which time Anne moves the toy to a box. When Sally returns, subjects are asked "Where will Sally look for her toy?" To successfully demonstrate a capacity for theory of mind, the subject needs to set aside their own knowledge of the toy's current location (in the box) and instead rely on Sally's representation of the location of the toy (in the basket). In other words, to successfully predict where Sally will look, subjects must represent Sally has having a false belief about where the toy is. Various adaptations of this type of task have been rig-

orously studied and led many to conclude that theory of mind requires explicit, effortful reasoning, and does not develop until around 3-5 years of age (Wellman, Cross, & Watson, 2001; Baron-Cohen et al., 1985).

Implicit theory of mind

Over the last decade or so, numerous tasks have been developed to measure theory of mind in pre-linguistic infants. In place of verbal responses to explicit questions, researchers have relied on non-verbal behavioral measures that may indicate the implicit calculation of other agents' mental states, measuring looking times in violation of expectation paradigms (Kovács et al., 2010; Onishi & Baillargeon, 2005), anticipatory looking when predicting where an agent will move (Schneider, Bayliss, Becker, & Dux, 2012; Surian & Geraci, 2012; Low & Watts, 2013; Southgate, Senju, & Csibra, 2007), or how infants decide to help others in interactive helping tasks (Buttelmann, Carpenter, & Tomasello, 2009).

In one such violation of expectation study, Onishi and Baillargeon (2005) introduced 15-month old infants to a Sally-Anne type paradigm. In the study, an experimenter was demonstrated to have a preference for an object. This object was then placed in one of two boxes. While the experimenter was not looking, this object then moved to the other location or remained in the same location as the experimenter last saw it. After the experimenter looks back at the boxes, she reaches for the ball either in the location where it was last seen or in the location where it actually was. This setup allowed the researchers to vary whether the experimenter's belief about the location of the ball was true or false. Onishi and Baillargeon (2005) found that infants looked longer when the experimenter searched for the ball in the location where it actually was than when the experimenter searched for the ball where she last saw it, suggesting that they expected the experimenter to act on the basis of her (now false) belief and were surprised when she didn't. Such non-linguistic behavioral measures have been argued to demonstrate that infants as young 15-months-old are capable of intention and belief attribution (Onishi & Baillargeon, 2005; Buttelmann et al., 2009; Träuble, Marinović, & Pauen, 2010; Kovács et al., 2010).

In addition to the work on implicit theory of mind representations in human infants, researchers have also sought to find evidence that human adults track others' mental states *automatically*, i.e., that they compute theory of mind representations mandatorily, even when it is counterproductive to the task being completed. As in the case of research on infant theory of mind, numerous paradigms have been used to measure automatic ToM in adults (Kovács et al., 2010; Phillips et al., 2015; Low & Watts, 2013; Schneider et al., 2012).

Promising results from work with human infants and studies on automatic theory of mind in adults has inspired researchers to claim that humans have an implicit theory of mind system, one that emerges early in development, does not require effortful control, and operates automatically (see, e.g., Apperly & Butterfill, 2009; Kovács et al., 2010).

Critically however, a debate continues over which kinds of theory of mind computations can be done implicitly. Some researchers have argued that the theory-of-mind capacity that arises early in development and operates automatically in adults allows one to compute others' false beliefs (e.g.., Kovács et al., 2010); others have instead argued that the evidence suggests that implicit theory of mind can only compute others' knowledge or ignorance, but not false beliefs (e.g., Phillips & Norby, 2019; Nagel, 2017; Surtees, Butterfill, & Apperly, 2012). The general distinction between these different kinds of mental states is one of *factivity* (Kiparsky & Kiparsky, 1970).

Factive and Non-Factive Theory of Mind

As previous discussed, many of the key results demonstrating implicit theory of mind have employed false-belief paradigms. Representations of beliefs are examples of non-factive mental states. These states require computing and reasoning about a representation that is strictly inconsistent with one's own understanding of the 'facts'; thus the term non-factive. While this kind of ToM representation has been the focus of the majority of ToM research, a number of researchers have argued non-factive tests neglect to capture more basic aspects of theory of mind, namely representations of *factive* mental states, e.g., what others' know, see, hear, are aware of, and so on (Martin & Santos, 2016; Nagel, 2017; Phillips & Norby, 2019).

In many of our pedestrian day-to-day interactions, we are primarily interested in keeping track of factive mental states. For example, you might want to remember your friend worked yesterday and has not seen the season finale of a show you both watch. Or you might have recently learned that your crush knows you own a cat and hope they bring it up in conversation. These representations are factive because they are directly tied to the way you take the world to actually be. Unlike our understanding of others' beliefs, which may or may not be false, we cannot represent others' as knowing or seeing things that we take to be false.

Consider the Queen of England. If you saw her at an event in Scotland, you can't represent someone else as *seeing* her in London, nor can you represent someone as *knowing* she's in Paris. However, you can, of course, represent someone else as *believing* that she is in London or Paris, or anywhere else for that matter. Importantly, factive mental states, like non-factive ones can still differ from your own: others' may *not* know something you do, or have seen something you have not, so they still often involve representing an understanding of the world that is different from one's own, even if it cannot be inconsistent with your own understanding.

Theory of mind researchers also employed implicit behavioral measurements such as reaction times in visual perspective-taking experiments in investigating implicit factive representations (Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014; Samson, Apperly, Braithwaite, Andrews, & Bodley-Scott, 2010). In one such task, Samson et al. (2010)

asked subjects to make judgements about the number of dots in a room that they or an avatar could see while manipulating the amount of time that participants had to respond. Crucially, the number of dots visible to the avatar was either identical to the number of dots visible to the subject not (i.e. the avatar only sees a subset of the number of dots visible to the subject). In the latter case, the avatar's perspective differs from one's own but is not inconsistent with it. Samson and colleagues found that when asked about their own perspective, subjects were slower to respond in the condition where the avatars perspective differed from their own than in the condition where it was was identical. Moreover, this was the case when subjects were explicitly instructed to only keep track of the number of dots that they themselves saw. These results were argued to show that subjects engaged in implicit factive ToM by automatically computing the avatar's perspective (Baker, Levin, & Saylor, 2016; Nielsen, Slade, Levy, & Holmes, 2015; Qureshi, Apperly, & Samson, 2010; Santiesteban et al., 2014).

Our contribution: We contribute to this theoretical debate over whether implicit theory of mind allows for the calculation of others' factive or non-factive mental states. To do so, we employ a novel meta-analytic approach. Specifically, we analyze the recent attempts at replicating the evidence for implicit theory of mind and ask whether we find a difference in the replicability of the results that originally provided evidence for factive vs. non-factive theory of mind.

Methods

In an ongoing meta-analysis, we have sampled 20 studies published between 2005 and 2018, containing 23 statistical tests that were taken to indicate the presence of factive ToM and 38 statistical tests that were taken to indicate the presence of non-factive ToM. These data represent 8 original paradigms and 12 replication studies including both developmental research and research on automatic theory of mind in adults (see Table 1).

Study Selection

Studies were selected by starting with the most heavily replicated paradigms. To be included, a study needed to either be the original publication of a paradigm or a direct replication. Papers were largely collected by conducting a literature review of direct replications of ToM tasks, and were expanded on through discussions with ToM researchers and a helpful recent meta-analysis on implicit theory of mind in infants (Barone, Corradi, & Gomila, 2019). For a complete list of the papers in each group see Appendix 1.

Many studies in the original literature review were subsequently excluded because they either have (a) not been replicated, (b) were conceptual replications not direct replications, or (c) did not involve tests of *implicit* theory of mind, which we defined as involving (i) studies with infants less than four-years-old (when explicit false belief tasks are passed) and (ii) studies that demonstrate automatic/mandatory theory of mind processing in adults. Future analyses may also want to broaden the scope of studies by relaxing these restrictions. However, for the purposes of this meta-analysis, we were only interested in the robustness of effect sizes across *direct replications* of factive and non-factive implicit theory of mind studies.

In addition, two of the original paradigms that satisfied our criteria had been empirically demonstrated to involve confounds, which once controlled for, no longer demonstrated the original effect that provided evidence for implicit theory of mind. Specifically, we were concerned about the confounds in Kovács et al. (2010) as demonstrated by Phillips et al. (2015) and the confounds in Buttelmann et al. (2009) as demonstrated by Priewasser, Rafetseder, Gargitter, and Perner (2018). Because of the concern that these original paradigms may not have measured implicit theory of mind in the first place, these pairs were excluded from the primary analyses (see Results for additional analyses that do include these paradigms).

Statistical Test Selection

Most papers selected contained multiple experiments. However, because we focus only on <Original, Direct Replication> pairs, we typically only include one or two experiments for each paper. For instance, original papers may introduce a paradigm in study one but then expand on those results in subsequent experiments by varying specific conditions or subject populations. If the first study of a paper was the only finding directly replicated in subsequent work, then only that experiment was included in our meta analysis. Similarly, if only a proper subset of the conditions of an original study were directly replicated (e.g., only conditions that did not include special populations), we excluded all conditions that had not been directly replicated from further analysis. In addition, replication research is often designed to replicate *and* extend original findings. However, because extensions of the original paradigms are not direct replications, these conditions were also set aside. Future research could consider relaxing some of these restrictions as well.

For each of the remaining experiments, the first author selected the statistical test that served as the key piece of evidence for theory of mind. In cases where this was not clear, the appropriate test was determined through a discussion between the first and second author. Similarly, each of these tests were classified as providing evidence for factive or non-factive theory of mind. Any disagreements were again resolved through discussion.

For replications, we attempted to match the critical statistic reported in the original paper. However, some replications did not report the same statistical tests as the original analyses. Additionally, some of the original analyses did not report effect sizes. In these instances, we either requested data from original authors or, in instances where open science practices made it possible, simply reanalyzed the data.

Finally, we converted all these effect sizes into a common standard, Pearson's *r*, which allowed us to compare across

paradigms that employed different statistical tests. Replication attempts with effects in the same direction as the originally reported effect were assigned positive effect sizes, while effects in the opposite direction were assigned negative effect sizes. The majority of the replication attempts were successful (19), albeit with typically smaller effect sizes than originally reported; however, 10 out of the 29 replication conditions were unsuccessful.

Coding

In addition to the common effect size, *r*, a number of other factors were recorded for each experiment (see below). It is our hope that this dataset will prove useful for others' in the field as well, and will be made avaialable upon completion.

1. Mean age and number of participants per condition
2. Paradigm: Eight levels of task type that describe which paradigm was used
3. Original or Replication: Two levels
4. Success or Failed Replication: Two levels
5. Factivitiy Classification: Two levels
6. Key Statistic: the critical statistical test that provides evidence for theory of mind.

Analysis Approach

To statistically investigate the replicability of evidence for factive and non-factive theory of mind, we compared a series of linear mixed-effects models to ask whether effect sizes were affected by (1) whether the test constituted the original demonstration of a theory of mind effect or was a replication, (2) whether the test demonstrated a capacity for factive or non-factive theory of mind, and (3) the interaction of these two two-level factors. While we originally hoped to use a more maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), our dataset is limited by the number of direct replication attempts that have currently been conducted, and we were not able to fit a complex random effects structure that included random slopes as well as random intercepts. Accordingly, our random effects structure only includes a random intercept for paradigm. As such, caution should be taken in generalizing the effects observed here to new paradigms. The significance of each reported effect was determined by comparing a model that included the relevant fixed-effect term (as well as other factors that were not currently being investigated) to a model that did not include that term (but did include the other factors not under investigation). The effect was determined to be significant if the fit of the model including the relevant term differed significantly from the fit of the model that did not include the relevant term.

Results

The analysis revealed no main effect of factivity ($\chi^2 = 0.132$, $p = 0.717$), but a significant effect of whether the study was an original or a replication ($\chi^2 = 7.181$, $p = 0.007$), which captured the fact that the original effect sizes were larger ($M = 0.475$, $SD = 0.171$) than the replication effect sizes ($M = 0.237$, $SD = 0.339$). Critically, however, this main effect was qualified by an interaction between whether the study was a replication and whether the paradigm tested for factive and non-factive ToM ($\chi^2 = 10.944$, $p < 0.001$) (see Figure 1). We explored this interaction effect by separately considering the effect of replication for factive and non-factive paradigms (see Table 1). For non-factive paradigms, we found that the replication effect sizes were significantly smaller ($M = .14$, $SD = .34$) than the original effect sizes ($M = .49$, $SD = .14$) ($\chi^2 = 16.589$, $p < 0.001$). For factive paradigms, by contrast, we found that the replication effect sizes were not significantly smaller ($M = .36$, $SD = .31$) than the original effect sizes ($M = .42$, $SD = .23$) ($\chi^2 = 0.139$, $p = 0.710$). Note that these conclusions are not substantively changed when we include the original and replication data from the Kovács et al. (2010) and Buttelmann et al. (2009) paradigms that were originally excluded because of concerns about possible confounds. We again find that for non-factive paradigms, the replication effect sizes were significantly smaller than the original effect sizes ($\chi^2 = 16.113$, $p < 0.001$), while there was no similar difference for factive paradigms ($\chi^2 = 0.873$, $p = 0.350$).

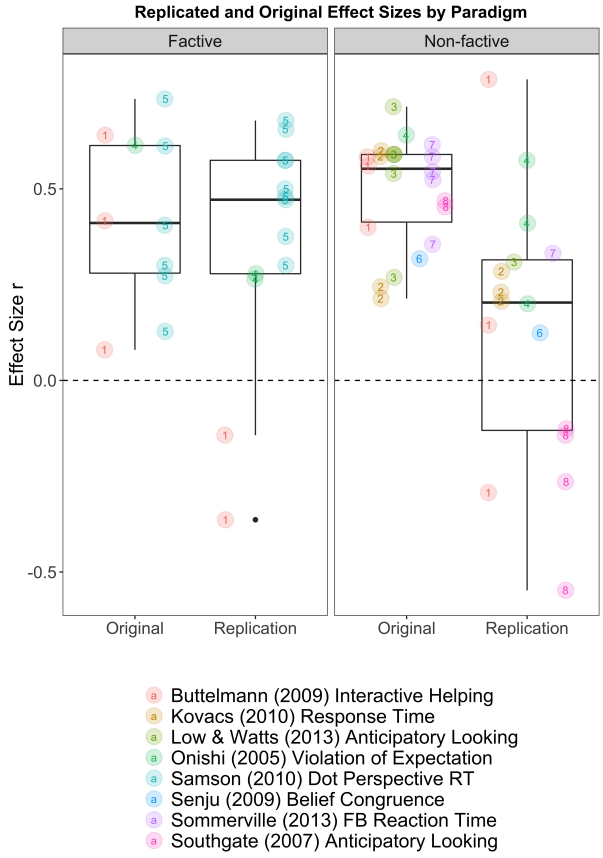

Figure 1: The bar chart shows the aggregate mean effect size for factive and non-factive ToM paradigms across original and replicated studies.

Factive			
Paradigm	Original	Replication	# of Tests
Buttelmann (2009) Interactive Helping	0.38	-0.25	5
Onishi (2005) Violation of Expectation	0.61	0.27	3
Samson (2010) Dot Perspective RT	0.41	0.51	15
Non-Factive			
Paradigm	Original	Replication	# of Tests
Buttelmann (2009) Interactive Helping	0.51	0.21	6
Kovacs (2010) Response Time	0.41	0.24	7
Low & Watts (2013) Anticipatory Looking	0.55	0.31	7
Onishi (2005) Violation of Expectation	0.64	0.39	4
Senju (2009) Belief Congruence	0.32	0.12	2
Sommerville (2013) FB Reaction Time	0.52	0.33	6
Southgate (2007) Anticipatory Looking	0.46	-0.27	6

Table 1: The table shows the aggregate mean effect size and number of data points for factive and non-factive ToM paradigms across original and replicated studies.

Discussion

Most research seeking to provide evidence for implicit theory of mind has primarily focused on demonstrating an ability to represent others' *non-factive* mental states, e.g., beliefs in a false belief paradigm (Barone et al., 2019). More recently, however, a number of researchers have argued representations of *factive* mental states, e.g., what others' know, see, hear, are more likely candidates for implicit ToM representations (in the sense that these computations appear earlier in phylogeny and ontogeny, are rendered more quickly, and require less effort) (Nagel, 2017; Phillips & Norby, 2019; Phillips, Knobe, Strickland, Armary, & Cushman, 2018).

In the ongoing meta-analysis reported on here, we asked whether the recent methodological push towards replication can be used as a tool that bears on the debate over the nature of implicit theory of mind: whether it more robustly represents factive or non-factive mental states. We found that, to the extent that there is replicable and robust evidence for implicit theory of mind, that evidence most clearly comes from tests that investigated *factive* theory of mind representations, rather than non-factive ones.

At the same time, it is important to note that the scope of the reported meta-analysis is limited by the small number of extant direct replication attempts, and the fact that many of these have focused on a few specific paradigms that were highly influential. Given our focus on direct replication, the data reported here are not nicely balanced: there are no cases of paradigms that involve original and replication effect sizes for both factive and non-factive tests across multiple behavioral measures. This is problematic because (as nicely demonstrated in a recent meta-analysis investigating implicit false belief tasks in human infants Barone et al. (2019)) the size of the observed effects can be dramatically impacted by the paradigm employed (e.g., anticipatory looking, violation of expectation). While such balanced data are unlikely to arise naturally, meta-analyses such as this one can be helpful in illuminating specific lacuna that may be addressed in future work. For example, we hope this work inspires researchers to employ more diverse measurements of factive theory of mind tasks. In the meantime, future meta-analytic work may aim to ameliorate these concerns somewhat by relaxing the inclusion criteria to allow for conceptual, as well as direct replications.

Theoretical replicability

While we were originally motivated to develop this meta-analytic approach out of our own interest in the theoretical distinction between factive and non-factive theory of mind (Phillips & Norby, 2019; Phillips et al., 2015), we hope that the current work inspires other researchers interested in a diversity of theoretical questions. We think of this approach as a proof of concept of the much broader potential for using the recent focus on replication to inform theoretically motivated questions in cognitive science. The most promising use-cases will be ones where cognitive scientists who have primarily been interested in theoretical questions pair with researchers who have primarily been focusing on conducting direct replications. In such cases, these research teams will be in a good position to pre-register the theoretically-motivated coding scheme, inclusion criteria, and analysis approach, putting them in an ideal position to have the meta-analytic results inform the theoretical question at hand. Finally, this project was made possible by researchers who spent valuable resources directly replicating existing paradigms and by researchers whose open data practices made it possible to reanalyze data in light of new theoretical distinctions. We hope that the current study further inspires researchers to continue this work, which we believe is critical for not only empirical but also theoretical progress to be made in cognitive science.

Appendix 1: Original vs. Replicated Papers

Factive

Originals Buttelmann et al. (2009); Onishi and Baillargeon (2005); Samson et al. (2010)

Replications Baker et al. (2016); Crivello and Poulin-Dubois (2018); Nielsen et al. (2015); Priewasser et al. (2018); Qureshi et al. (2010); Santiesteban et al. (2014); Träuble et al. (2010); Yott and Poulin-Dubois (2016)

Non-Factive

Originals Buttelmann et al. (2009); Kovács et al. (2010); Low and Watts (2013); Onishi and Baillargeon (2005); Schneider et al. (2012); Senju, Southgate, White, and Frith (2009); Sommerville, Bernstein, and Meltzoff (2013)

Replications Bernstein, Thornton, and Sommerville (2011); Crivello and Poulin-Dubois (2018); Kulke, von Duhn, Schneider, and Rakoczy (2018); Phillips et al. (2015); Priewasser et al. (2018); Träuble et al. (2010); Yott and Poulin-Dubois (2016, 2012)

References

Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116(4)*, 953.

Baker, L., Levin, D., & Saylor, M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance.*, *42*, 508-516.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *1*, 37-46.

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior and Development*, *57*, 101350.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bennett, J. (1978). Some remarks about concepts. *Behavioral & Brain Sciences*, *1*, 557–560.

Bernstein, D., Thornton, W., & Sommerville, J. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research.*, *37*, 481-502.

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.

Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development*, *46*, 51-57.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral & Brain Sciences*, *1*, 568.

Kiparsky, P., & Kiparsky, C. (1970). Fact. In M. Bierwisch & K. E. Heidolph (Eds.), *Progress in linguistics: a collection of papers.* Walter de Gruyter GmbH & Co. KG.

Kovács, A. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830–34.

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? results from a systematic replication study. *Psychological science*, *29(6)*, 888-900.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24(3)*, 305-311.

Martin, A., & Santos, L. (2016). What cognitive representations support primate theory of mind?. *Trends in cognitive sciences*, *20(5)*, 375-382.

Nagel, J. (2017). Factive and nonfactive mental state attribution. *Mind & Language*, *32(5)*, 525-544.

Nielsen, M., Slade, L., Levy, J., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process?. *The Quarterly Journal of Experimental Psychology.*, *68*, 1931-1951.

Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308(5719)*, 255-258.

Phillips, J., Knobe, J., Strickland, B., Armary, P., & Cushman, F. (2018). Evidence for evaluations of knowledge prior to belief. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 40).

Phillips, J., & Norby, A. (2019). Factive theory of mind. *Mind & Language*.

Phillips, J., Ong, D., Surtees, A., Xin, Y., Williams, S., Saxe, R., & Frank, M. (2015). A second look at automatic theory of mind: Reconsidering kovács, téglás, and endress (2010). *Psychological Science*, *26(9)*, 1353-1367.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(4), 515–526.

Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2018). Helping as an early indicator of a theory of mind: Mentalism or teleology?. *Cognitive Development*, *46*, 69-78.

Qureshi, A., Apperly, I., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117(2)*, 230-236.

Samson, D., Apperly, I., Braithwaite, J., Andrews, B., & Bodley-Scott, S. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1255-1266.

Santiesteban, I., Catmur, C., Hopkins, S., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing?. *Journal of Experimental Psychology: Human Perception and Performance*, *40(3)*, 929.

Schneider, D., Bayliss, A., Becker, S., & Dux, P. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of experimental psychology: general*, *141(3*, 433.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in asperger syndrome. *Science*, *325(5942)*, 883-885.

Sommerville, J., Bernstein, D., & Meltzoff, A. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development.*, *84*, 1846-1854.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18(7)*, 587-592.

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130(1)*, 1-10.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, *30(1)*, 30-44.

Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking

in children and adults. *British Journal of Developmental Psychology*, *30*(1), 75–86.

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, *15(4)*, 434-444.

van der Wel, R., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? evidence from a continuous measure. *Cognition*, *130*, 128–133.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72(3)*, 655-684.

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *British Journal of Developmental Psychology*, *30(1)*, 156-171.

Yott, J., & Poulin-Dubois, D. (2016). Are infants' theory-of-mind abilities well integrated? implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development*, *17(5)*, 683-698.