Cognition 136 (2015) 49–63

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Specialized mechanisms for theory of mind: Are mental representations special because they are mental or because they are representations?

Adam S. Cohen [a,b,*], Joni Y. Sasaki [c], Tamsin C. German [d]

[a] Department of Psychology, University of Western Ontario, London, ON N6A 3K7, Canada
[b] The Brain and Mind Institute, University of Western Ontario, London, ON N6A 5B7, Canada
[c] Department of Psychology, York University, Toronto, ON M3J 1P3, Canada
[d] Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Received 30 December 2012
Revised 30 September 2014
Accepted 17 November 2014

Keywords:
Theory of mind
Domain-specific
Functional specialization
Belief reasoning
Linguistic representation
Metarepresentation

ABSTRACT

Does theory of mind depend on a capacity to reason about representations generally or on mechanisms selective for the processing of mental state representations? In four experiments, participants reasoned about beliefs (mental representations) and notes (non-mental, linguistic representations), which according to two prominent theories are closely matched representations because both are represented propositionally. Reaction times were faster and accuracies higher when participants endorsed or rejected statements about false beliefs than about false notes (Experiment 1), even when statements emphasized representational format (Experiment 2), which should have favored the activation of representation concepts. Experiments 3 and 4 ruled out a counterhypothesis that differences in task demands were responsible for the advantage in belief processing. These results demonstrate for the first time that understanding of mental and linguistic representations can be dissociated even though both may carry propositional content, supporting the theory that mechanisms governing theory of mind reasoning are narrowly specialized to process mental states, not representations more broadly. Extending this theory, we discuss whether less efficient processing of non-mental representations may be a by-product of mechanisms specialized for processing mental states.

1. Introduction

Theory of mind, the ability to predict and explain behavior in terms of mental states, has been described from many angles. It develops similarly across cultures (Avis & Harris, 1991; Barrett et al., 2013; Callaghan et al., 2005; Liu, Wellman, Tardif, & Sabbagh, 2008), emerges early in human development (Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007), and appears in related forms in non-human primates (Call & Tomasello, 2008). Especially striking are cases of "mindblindness" in which the ability to "see" or represent the mental states of others is impaired, leading to profound social difficulties that may be diagnostic of autism and other conditions (Baron-Cohen, Leslie, & Frith, 1985; Frith, 1992). In addition to these empirical discoveries, it has been suggested on logical grounds that theory of mind is a ubiquitous, indispensable, and largely effortless part of human reasoning (Dennett, 1987; Fodor, 1987).

Beyond describing theory of mind, researchers from a range of disciplines have spent the last three decades

* Corresponding author at: Department of Psychology, University of Western Ontario, London, ON N6A 3K7, Canada. Tel.: +1 519 661 2111x84596.

E-mail address: acohen42@uwo.ca (A.S. Cohen).

http://dx.doi.org/10.1016/j.cognition.2014.11.016

exploring how the underlying neurocognitive mechanisms work (for reviews, see Frith & Frith, 2006; Goldman, 2006; Harris, 2006; Saxe, Carey, & Kanwisher, 2004). To illuminate the properties of these mechanisms, researchers have relied heavily on different versions of "false-belief" tasks (Baron-Cohen et al., 1985; Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983). Consider the following commonly used false-belief scenario: *Before Sally leaves for lunch, she hides her ball in the basket. While she is away eating, her big sister Anne plays a trick on her and moves her ball from the basket to the box. When Sally returns, where will she look for her ball?* To succeed on the task, the participant needs to attribute a belief to Sally that she (falsely) believes the ball is in the basket and, on the basis of that false belief, will search in the basket, not in the box where the ball is really located and where the participant actually knows it to be. While many find this task trivial, not everyone passes: children who are 3 years old or younger (Wellman, Cross, & Watson, 2001) and children with autism (Baron-Cohen et al., 1985) fail to attribute a false belief, answering that Sally will search where the object really is, not where she thinks it is. Explaining these developmental and neuropsychological findings turns on understanding how the underlying neurocognitive mechanisms work.

Despite the intense research effort, considerable disagreement surrounds the properties of the mechanisms supporting mental state attribution. An especially contentious issue concerns whether these mechanisms handle a broad class of representations (Corballis, 2003; Doherty, 2000; Iao & Leekam, 2014; Iao, Leekam, Perner, & McConachie, 2011; Leekam & Perner, 1991; Leekam, Perner, Healey, & Sewell, 2008; Perner, 1991, 1995; Stone & Gerrans, 2006; Suddendorf, 1999), or whether they are specialized for processing mental state representations (Baron-Cohen, 1995; Cohen & German, 2010; Frith & Frith, 2003; Leslie, 1994; Leslie, Friedman, & German, 2004; Leslie & Thaiss, 1992; Saxe et al., 2004; Scholl & Leslie, 1999; Scott & Baillargeon, 2009; Sperber, 2000). These two accounts, the representational processing theory and the mental state processing theory, both propose specialization to some extent, but disagree about scope: is there a broadly specialized capacity for understanding representations or a narrowly specialized mechanism for processing mental representations?[1] The current investigation tests these two competing theories by comparing adults' ability to reason about beliefs (mental representations) and written notes (non-mental, linguistic representations). The comparison between mental and linguistic representations addresses a triplet of problems that have not been simultaneously confronted in previous work. Before addressing these three problems more fully, we discuss the two theories under consideration.

1.1. Broad specialization: mechanisms for understanding representation

One influential account of theory of mind, the representational processing theory, proposes that mental states are understood not according to principles specific to the mental domain but as part of the larger domain of representation (e.g., Perner, 1991; Suddendorf, 1999). Pre-empirically, why suppose a capacity for processing representations of representations, sometimes described as a *capacity to metarepresent*, which treats belief, intention, and other mental states the same as maps, signs, and other non-mental representations? Philosophical considerations suggest that mental and non-mental representations (a) refer to things and (b) refer to things as being a certain way (Dretske, 1988; Fodor, 2008; Goodman, 1976). For instance, the belief that Sandy Koufax played for the Los Angeles Dodgers and the belief that he played for the Brooklyn Dodgers both refer to Sandy Koufax, but critically they represent him in different ways – as being on two different teams. In the non-mental domain, if two baseball cards picture Sandy Koufax in different uniforms, then like the belief example, the cards both refer to Koufax but differ in how he is represented. According to the representational processing theory, grasping the difference between *what* is represented (e.g., Sandy Koufax) and *how* it is represented (e.g., as a Los Angeles or Brooklyn Dodger) is the critical conceptual capacity underlying an understanding of both mental and non-mental representation.[2]

1.2. Narrow specialization: mechanisms for understanding mental states

Alternatively, theory of mind may depend on a collection of mechanisms, at least one of which is specialized for processing mental states (Baron-Cohen, 1995; Leslie & Thaiss, 1992; Leslie et al., 2004). On this theory, a specialized mechanism operates over a restricted class of inputs including eye gaze (Baron-Cohen, 1995), contingent motion (Johnson, 2003), self-propelled movement (Leslie, 1994; Premack, 1990), and other social cues; integrates these cues; and then applies specialized procedures to compute a set of candidate mental states. Successful reasoning according to this account depends on the operation of additional mechanisms, including a selection processor that selects one of the candidate mental states, enabling conscious and unconscious prediction and explanation of an agent's behavior in mentalistic terms (see e.g., Leslie & Polizzi, 1998; Leslie & Thaiss, 1992). Executive selection processes and other mechanisms required for theory of mind performance (e.g., working memory, perceptual

[1] We resist calling the mechanisms posited by the representational processing theory "domain-general" or "general-purpose" because compared to most content-free, domain-general accounts (e.g., statistical learning mechanisms; Turk-Browne, Junge, & Scholl, 2005), this theory proposes processing over representations of representations (i.e., metarepresentations), a limited class of input. This at a minimum commits the theory to some degree of domain-specificity and probably functional specialization.

[2] This is the sense vs. reference (Dretske, 1988; Frege, 1892/1980) or representing vs. representing-as (Fodor, 2008; Goodman, 1976) distinction, except on the representational processing theory, it is not just a philosophical distinction; it is also one that ordinary people make (Perner, 1991; Perner, Mauer, & Hildenbrand, 2011). An even finer philosophical distinction could be drawn (e.g., between properties of the referent and modes of presentation) producing different types of what–how distinctions; however, conflating these is not a problem for the representational processing theory if it is assumed these are philosophical distinctions and not psychological distinctions in the minds of everyday people.

systems, etc.), are likely recruited for other tasks outside the theory of mind domain, including tasks involving non-mental representations. In contrast, specialized mechanisms involved in theory of mind computation, because of their domain-specific structure, do not extend their processing to notes, photos, signs, or other non-mental representations. The theory assumes that a separate, possibly (though not necessarily) less-specialized device handles non-mental representations, suggesting that these two systems are in principle dissociable.

1.3. Testing broad and narrow specialization theories

In theory of mind research, tests of broad and narrow specialization theories have often relied on mental and non-mental representation reasoning tasks that are matched in all respects except for the type of representation that is introduced. If differences in performance are observed, this isolates the cause to the manipulation of representation type and strongly suggests the underlying systems *discriminate between the representations*, as expected on the narrow specialization account (Leslie & Thaiss, 1992). In contrast, given a system with the less specialized capacity to understand and process representations, changing representation type should have little effect on performance because for that kind of system, *representation types are not distinguished; they are "psychologically" (i.e., computationally) equivalent* (Iao et al., 2011). A key distinction between the two theories, then, is whether processing of mental and non-mental representations is expected to dissociate, as predicted on the mental state processing theory, or associate, as predicted on the representational processing theory.

There are multiple sources of evidence for the dissociations predicted on the narrow specialization theory. One source of evidence stems from neuropsychological dissociations, including selective impairment on false-belief tasks despite intact performance on closely matched "false-photograph" tasks (Zaitchik, 1990) in children with autism (Leekam & Perner, 1991; Leslie & Thaiss, 1992). Neurovascular dissociations observed in fMRI studies with healthy adults further support this theory, revealing a specialized network in the brain including bilateral temporoparietal junction, precuneus, and medial prefrontal cortex that responds more strongly when engaged in processing beliefs than photos (Saxe & Kanwisher, 2003). Based on the reasoning above, these neuropsychological and neurovascular dissociations provide evidence for mental state specialization and furthermore are not predicted on the representational processing account.

1.4. The falseness problem

Although the false-photograph task used in previous investigations has significant appeal because of the tight control it introduces by closely matching the false-belief task, it suffers from "the falseness problem:" photos, assuming they were not tampered with, are not necessarily false when the world changes because they continue to refer to and accurately represent the past, making them true representations of past events (Leekam & Perner, 1991; Perner & Leekam, 2008). Beliefs, on the other hand, typically refer to the present and can be false if the world changes in ways inconsistent with the belief. For example, a vacation photo of a friend sitting on the beach, unlike a belief with the same content, remains true if the friend is currently working at the office. The difference then is that photos refer to how the world *was*; beliefs refer to how the world *is*.

In light of these considerations, researchers have questioned the validity of using false photos as a control for false belief, arguing that false beliefs should be computationally more demanding because the mind has to reconcile a discrepancy between someone's belief about the world (*Sally believes* that the ball is in location A based on seeing it there at time 1), and the current state of the world (*the fact* that the ball was moved and is in location B at time 2), putting extra load on executive functioning to handle the incongruency. With photos, even if the world changes (*the fact* that the ball is now in location B), there is no conflict between the photo and the state of the world it refers to since the photo continues to correctly represent the world the way it was (*the photo shows* the ball, as it was, in location A; e.g., Sabbagh, Moses, & Shiverick, 2006). Given this difference in "falseness," the dissociations initially observed in children with autism are harder to interpret: rather than a domain-specific impairment, the false-belief task may simply be more executively demanding than the photo task. Similarly, differences in brain activity (Saxe & Kanwisher, 2003) could merely reflect greater activation of networks involved in executive function when engaged in false-belief reasoning. Supporting these suggestions, research has shown that executive function ability predicts performance on false-belief but not false-photo tasks in typically developing children, suggesting that belief and photos make different demands on executive function (Sabbagh et al., 2006).

Because the conceptual problem that photos introduce is specific to representations that refer to the past, substituting in non-mental representations that, like beliefs, refer to the present, can resolve the falseness problem. Other pictorial representations such as maps and signs meet this requirement: they function by referring to the here-and-now after being created, and therefore can be genuinely false if, for example, the world changes. Two approaches have been taken to address the falseness problem: the association approach and the dissociation approach.

1.5. Resolving the falseness problem I: the association approach

Among typically developing children, correlations establish false-belief tasks as structurally and conceptually more similar to false-sign than false-photo tasks (Iao & Leekam, 2014; Leekam et al., 2008). Additionally, children with autism perform similarly with verbal and non-verbal false-belief and false-sign tasks (Bowler, Briskman, Gurvidi, & Fornells-Ambrojo, 2005; Iao & Leekam, 2014). These findings in children with and without autism challenge the interpretation of dissociations between beliefs and photos as signaling specialization and suggest a common capacity to represent mental and other representations, as predicted on the broad representation processing theory. Neuroimaging evidence further supports this view, revealing shared processing of beliefs and signs in the left temporoparietal junction (Aichhorn et al., 2009; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006).

1.5.1. The association problem

Although these studies have many virtues and avoid the problems plaguing false photos, they are difficult to interpret because of "the association problem:" both the representational processing theory and the mental state processing theory predict associations between these tasks because of 3rd variables. The representational processing account predicts a correlation between performance on theory of mind and certain non-mental representation tasks because of common conceptual demands to process representations. The mental state processing account expects an association because of common performance demands. In particular, because "there is more to passing the false-belief task than theory of mind" (Bloom & German, 2000), including demands on attention, memory, and executive function (e.g., inhibiting prepotent responses and selecting belief contents; Leslie, German, & Polizzi, 2005; Leslie & Polizzi, 1998), and given that the false-belief and false-sign tasks are nearly identically structured, performance may be linked because of these other factors.

Aside from performance factors, correlations could emerge if distinct systems for processing mental and non-mental representations develop along a similar timetable. Just as one might expect a correlation between shoe size and math ability among preschoolers for reasons entirely unrelated to common conceptual development, the construction of two computationally independent mechanisms might be yoked because of a third developmental factor relevant to both systems, driving similar schedules of development and producing correlations in performance. Possible computational factors include availability of information in the developmental environment relevant to programming both systems, and potential neurodevelopmental factors include increased myelination or pruning that might influence how both systems develop.

The association problem leads to difficulty adjudicating between multiple, equally plausible explanations for why processing of mental and non-mental representation are associated[3]; therefore, other research approaches may provide a clearer test between the broad and narrow specialization theories.

1.6. Resolving the falseness problem II: the dissociation approach

As an alternative to testing for associations, researchers have also tested for dissociations using pictorial representations other than photos. By taking a dissociation approach with closely matched tasks that factor in the falseness problem, a number of studies have demonstrated that children with autism show striking dissociations in performance, marked by selectively impaired reasoning about false beliefs but intact reasoning about false maps (Leslie & Thaiss, 1992), false drawings (Charman & Baron-Cohen, 1992), and false models (Charman & Baron-Cohen, 1995). More recently, in a series of experiments with healthy adults using sensitive reaction time measures, belief processing was significantly faster than processing of maps and signs (Cohen & German, 2010). These findings, which are based on studies that resolve the association and falseness problems, provide evidence for the mental state processing theory and are not predicted on the representational processing theory.

1.6.1. The content problem

Experiments using maps, drawings, models, and signs are nonetheless challenged by "the content problem." Representational and mental state reasoning theories assume representations of belief have propositional (sentence-like, logical) content (Leslie & Thaiss, 1992; Perner, 1995), but research has relied on comparing them to non-mental representations that are generally assumed to have pictorial (picture-like, iconic) content. These types of representations and their associated content potentially differ from each other on a crucial dimension. While propositional content can be true or false, exhibiting "truth-evaluability," pictorial content is generally thought to lack this property, having accuracy conditions (i.e., they can be more or less accurate) rather than straightforward truth conditions (Crane, 2009; Lopes, 1996, 2003; for a dissenting view, see Perini, 2005). Although this claim may be contentious in philosophy, whether it is secure or not is irrelevant to our research question since the present research examines not the nature of representation, but the psychology of representation. Our main goal is to motivate a specific psychological hypothesis derived from philosophical considerations.

Differences in the properties of mental and pictorial representations suggest an alternate hypothesis for previously observed dissociations on otherwise matched mental and pictorial reasoning tasks. On this counterhypothesis, processing advantages for belief may emerge because the underlying mechanisms are specialized for processing propositional rather than mental representations. One way a propositional account could explain previous results without appealing to mental state specialization is if the mind has to convert pictorial representations into a propositional code and impose truth-values on them in order to process them, which would introduce costs in processing accuracy or speed. On this account, evidence that beliefs are processed faster may arise from the computational costs of converting pictorial representations, but not beliefs, into propositional representations. Privileged processing of propositional representations, while not expected on the mental state processing theory, is consistent with the representational processing theory, which, in one of its more prominent formulations, stipulates that children acquire a theory that allows them to evaluate a

[3] Associations predicted on the representational processing theory might be easier to interpret if the tasks employed shared little in common except the hypothesized common demand to metarepresent, permitting one to locate any observed correlation with the unique factor of metarepresentation shared between tasks. However, this approach is rarely taken (but see Doherty, 2000; Doherty & Perner, 1998; Garnham, Brooks, Garnham, & Ostenfeld, 2000 for exceptions).

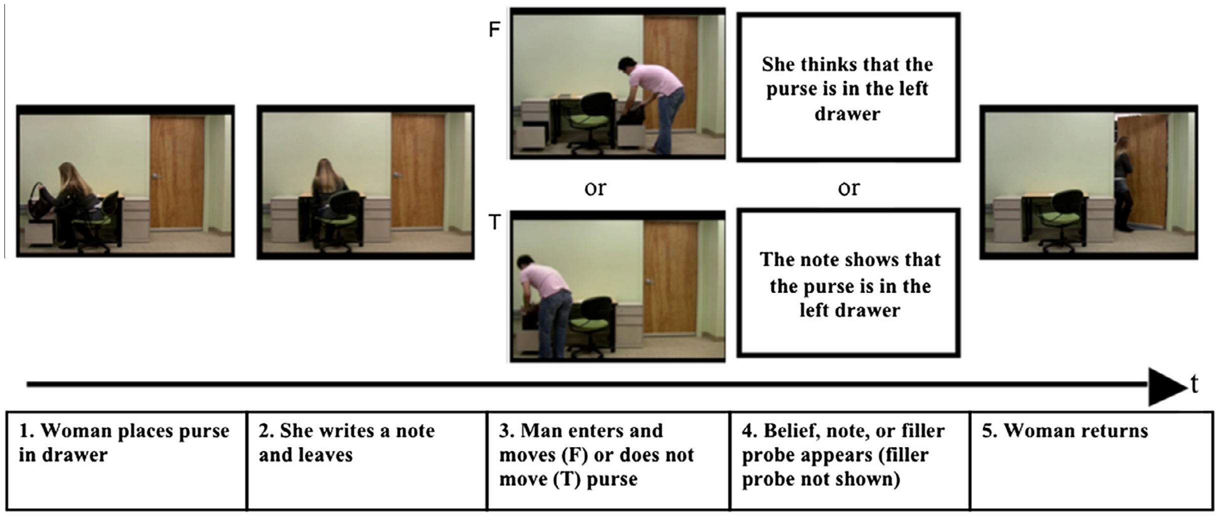

Fig. 1. The hybrid note vs. belief task (F = false condition and T = true condition). See text for details.

proposition as true or false and to consider "a different truth value than the one it has," such as when an agent has a false belief (Perner, 1995).[4]

1.6.2. Resolving the content problem: beliefs, notes, and propositional content

To address the content problem, we suggest comparing linguistic and mental representations. Linguistic representations offer a critical advantage. Written notes are representationally closer to beliefs than either maps or signs because mental and linguistic representations stand in relation to *propositions*. Propositions can be the objects of attitudes like belief or expressed by sentences like those written in a note. This representational closeness makes for an especially strong test of the broad vs. narrow specialization theories and is the focus of the four studies presented here.

1.7. The current experiments: belief vs. note task

In four experiments, we used a hybrid false-belief and false-note task with neurotypical adults to address the triplet of problems. To address the falseness problem and avert the issues encountered with photos, we (a) used a written note (i.e., non-mental representation) that, like a belief, could be false and (b) included "true" conditions, which required reasoning about either true beliefs or true notes. In response to the association problem, we took a dissociation approach and used a task that closely matched the structure of the false-belief task but introduced a non-mental representation. To resolve the content problem and the issues raised with maps and signs, we used linguistic representations that expressed propositions.

Several other features were incorporated into the belief vs. note task. To add further control, the two tasks were merged into a single scenario (see Fig. 1), ensuring that all events leading up to and following responses were identical. Therefore, it was impossible to discriminate the belief condition from the note condition based on anything other than the actual test probe, which required participants to infer either the content of a person's belief or the content of a note. To obtain a more sensitive measure than the accuracy scoring used in preschool studies, we measured both reaction times and accuracy in adults.

Experiment 1 used a format-unstressed task in which representational properties of the note were not made particularly salient. In Experiment 2, a format-stressed task was introduced to emphasize the format of the representation. The logic for using these different tasks was that making representational properties explicit might favor triggering representational knowledge, which the representational processing theory considers critical to understanding mental and non-mental representation. Experiments 3 and 4 replicated Experiments 1 and 2, respectively, ruling out an alternative explanation that the task imposed greater demands for reasoning about notes than beliefs.

The first hypothesis, derived from the representational processing theory, predicts that accuracies and reaction times to belief and note probes should be similar since they both depend on the same underlying capacity for processing propositional representations. In contrast, the second hypothesis, derived from the mental state processing theory, predicts that reaction times should be significantly faster, accuracies significantly higher, or both for belief compared to note probes, reflecting more efficient, specialized processing of mental representations.

[4] The falseness and content problems may be related: they both propose that what makes a belief true or false may not necessarily apply to certain pictorial representations. While photos and beliefs seem to have different falsity-conditions in false-representation tasks (the falseness problem), maps and signs compared to beliefs seem to differ in truth-evaluability (the content problem). Both challenge the validity of pictorial representations as controls for testing specialization.

2. Experiment 1

2.1. Method

2.1.1. Participants

Fifty-five participants, all undergraduates participating for course credit, were tested. Out of the 55 participants, seven participants were not analyzed, either for failing to meet accuracy criteria ($n = 3$)[5] or for completing too few trials due to a computer glitch ($n = 4$).[6] Thus, the final sample included 48 participants (32 females and 15 males; $M_{age} = 18.74$, $SD_{age} = 1.08$).[7]

2.1.2. Design and procedure

In a 2 (Representation: mental vs. linguistic) × 2 (Truth-value: true or false) repeated-measures design, participants received 12 trials in each condition and an additional 44 filler trials for a total of 92 trials. The filler trials were randomly inserted throughout the session to keep trial order unpredictable and to prevent participants from preparing responses.

Participants watched videos of simple search–action scenarios on a computer (see selected frames in Fig. 1). During the practice round prior to the test phase, participants were given three practice trials to familiarize them to the task and ensure they understood the scenarios. On the first practice trial, the video paused at various points to allow participants time to read subtitles that conveyed the cover story (see Appendices A and B for complete instructions and text of subtitles). In each video, an actress put a purse belonging to a friend in a drawer and, before leaving the room, reached for a piece of paper and wrote a note for the friend indicating the purse's location. While gone, an officemate stopped by to put away an unrelated object and either moved the purse to the opposite drawer ("false" condition) or left it in place ("true" condition). At this point, the video was interrupted with a belief or note test probe (e.g., "The girl thinks the purse is in the right drawer"; "The note shows the purse is in the right drawer") or one of several kinds of filler probes (e.g., "It is true that the purse has switched drawers"; "It is true that the girl wrote a note"; "It is true that the purse is in the left drawer"; "It is true that the book is red"). Participants were instructed to provide "yes" or "no" responses to probes and to respond as quickly and as accurately as possible. The "y" and "n" keys were used for "yes" and "no" responses, respectively. The participants used their dominant hand, which they were instructed to keep over the "h" key (which sits between the "y" key above it and the "n" key below it). On practice trials, filler probes were used and feedback was provided informing participants whether their button press response to the probe was correct or incorrect. No feedback was provided during the test phase. After the button press response, the video finished and as part of a distractor task, participants gave a pointing response to indicate the object's actual location, ensuring they followed instructions and tracked the object.

Videos ranged from 23 to 25 s and were presented with E-Prime 2.0 software (Psychology Software Tools, Inc.). To maximize the number of trials, video playback speed was increased to 3 times the recorded speed. For all four conditions, responses to probes were equally likely to be "yes" or "no" and the object was equally likely to (a) start off on the left or right side and (b) end up on the left or right side at the time of the probe.

2.2. Results and discussion

In an accuracy analysis, percent correct, determined by dividing number of correct trials by total trials in each condition, was entered into a two-way repeated-measures analysis of variance (ANOVA) and revealed a main effect of representation type, $F(1,47) = 3.95$, $p = .053$, $\eta_p^2 = .078$, indicating percent correct was significantly higher for belief than note probes, and a main effect of truth value, $F(1,47) = 9.97$, $p = .003$, $\eta_p^2 = .175$, indicating that percent correct was significantly higher for true than false representations. These main effects were qualified by a marginally significant representation-type by truth-value interaction, $F(1,47) = 3.07$, $p = .086$, $\eta_p^2 = .061$. The means and standard errors are plotted in Fig. 2.

Pairwise post hoc t-tests confirmed that the accuracy comparison between false beliefs ($M = 93.2\%$, $SD = 0.12$) and false notes ($M = 88.8\%$, $SD = 0.19$) was significant, $t(47) = 2.12$, $p = .039$, $d = 0.306$, although there was no difference between true beliefs ($M = 95.3\%$, $SD = 0.12$) and true notes ($M = 94.8\%$, $SD = 0.12$), $t(47) = 0.48$, $p = .635$, $d = 0.069$. There was also a significant difference between false and true notes, $t(47) = 2.81$, $p = .007$, $d = 0.405$, and a marginal difference between false and true beliefs, $t(47) = 1.86$, $p = .069$, $d = 0.268$.

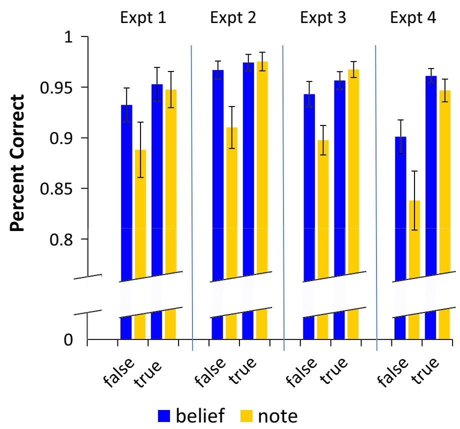

Fig. 2. Accuracy as a function of representation type (belief vs. note) and truth value (true vs. false) in the format-unstressed task (Expt 1), format-stressed task (Expt 2), modified format-unstressed task (Expt 3) and modified format-stressed task (Expt 4). Error bars represent SEM.

---

[5] An overall accuracy of 62.5% correct was used as a cutoff because it marked the point at which performance was statistically better than chance.

[6] These participants received fewer than 30 out of 48 trials. Including these subjects in the analyses had no effect on the results and, if anything, tended toward strengthening the reported patterns.

[7] Participant information (age and gender) was missing for one person included in the analysis.

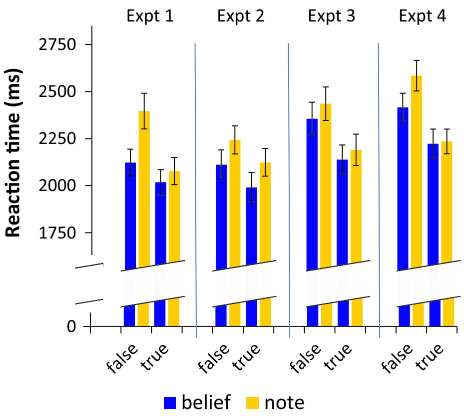

Fig. 3. Reaction times as a function of representation type (belief vs. note) and truth value (true vs. false) in the format-unstressed task (Expt 1), format-stressed task (Expt 2), modified format-unstressed task (Expt 3) and modified format-stressed task (Expt 4). Error bars represent SEM.

In a response time analysis, reaction times were subjected to a two-factor repeated-measures ANOVA,[8] revealing a main effect of representation type, $F(1,47) = 26.3$, $p < .001$, $\eta_p^2 = .359$, such that reaction times to beliefs were significantly faster than reaction times to notes, a main effect of truth value, $F(1,47) = 37.2$, $p < .001$, $\eta_p^2 = .442$, such that reaction times were faster to true than false representations, as well as a representation-type by truth-value interaction, $F(1,47) = 10.7$, $p = .002$, $\eta_p^2 = .186$. Inspection of Fig. 3 suggests that the interaction stems from the difference between beliefs and notes being larger for false representations than for true representations.

Pairwise post hoc *t*-tests confirmed that the reaction time comparison between false beliefs ($M = 2123$ ms, $SD = 494$ ms) and false notes ($M = 2397$ ms, $SD = 657$ ms) was significant, $t(47) = 4.92$, $p < .001$, with an effect size ($d = 0.710$) larger than that obtained between true beliefs ($M = 2019$ ms, $SD = 464$ ms) and true notes ($M = 2078$ ms, $SD = 500$ ms), $t(47) = 1.74$, $p = .088$, $d = 0.251$. There were also differences between false and true notes, $t(47) = 5.88$, $p < .001$, $d = 0.849$, and false and true beliefs $t(47) = 2.60$, $p = .013$, $d = 0.375$.

The results of Experiment 1 suggest that probing for information conveyed as a public linguistic representation (i.e., "The note shows the purse is in the left drawer") did not result in processing that was as efficient as for mental states. Not only were participants more accurate on false-belief trials, they were also faster. As suspected, effects were more pronounced in the reaction time than in the accuracy analysis. Although the difference between true-belief and true-note conditions did not reach significance, we tentatively suggest that reaction times were approaching a performance floor.

One possible objection to Experiment 1 is that in order to activate the metarepresentational capacity, a representation must make clear that it has representational properties. The probe about the note might have failed to fully trigger the metarepresentational capacity because it did not stress any of those properties. To address this counter-explanation, a second experiment was run in which medium, a property of the representation, was stressed in the probe.

3. Experiment 2

3.1. Method

3.1.1. Participants

Fifty-three participants, all undergraduates participating for course credit, were tested. Of the 53 participants, one failed to meet accuracy criteria and was not analyzed, leaving the final sample at $N = 52$ (40 females and 10 males; $M_{age} = 18.84$ years, $SD_{age} = 0.89$ years).[9]

3.1.2. Design and procedure

The design and procedure were identical to Experiment 1 except that the belief and note probes were replaced with "In her mind, the purse is in the right drawer" and "In the note, the purse is in the right drawer," respectively, in order to highlight format.

3.2. Results and discussion

In an accuracy analysis, percent correct was entered into a two-way repeated-measures analysis of variance (ANOVA) and revealed a main effect of representation type, $F(1,51) = 6.04$, $p = .017$, $\eta_p^2 = .106$, indicating percent correct was significantly higher for beliefs than notes, and a main effect of truth value, $F(1,51) = 9.47$, $p = .003$, $\eta_p^2 = .157$, indicating that percent correct was significantly higher for true than false representations. These main effects were qualified by a representation-type by truth-value interaction, $F(1,51) = 7.59$, $p = .008$, $\eta_p^2 = .130$. The means and standard errors are plotted in Fig. 2.

Pairwise post hoc *t*-tests confirmed that the accuracy comparison between false beliefs ($M = 96.7\%$, $SD = 0.06$) and false notes ($M = 91.0\%$, $SD = 0.15$) was significant, $t(51) = 2.92$, $p = .005$, $d = 0.405$, although there was no difference between true beliefs ($M = 97.4\%$, $SD = 0.06$) and true notes ($M = 97.5\%$, $SD = 0.07$), $t(51) = 0.097$, $p = .923$, $d = 0.013$. There was also a significant difference between false and true notes, $t(51) = 3.36$, $p = .001$, $d = 0.466$, but no difference between false and true beliefs, $t(51) = 0.663$, $p = .510$, $d = 0.092$.

Reaction times were subjected to a two-factor repeated-measures ANOVA, revealing a main effect of representation type, $F(1,51) = 9.49$, $p = .003$, $\eta_p^2 = .157$, such that reaction times to beliefs were significantly faster than reaction time to notes, and a main effect of truth value, $F(1,51) = 15.8$, $p < .001$, $\eta_p^2 = .237$, such that reaction times were significantly faster to true than false representations. The representation-type by truth-value interaction was not

[8] For all analyses, outlier trials (defined as ±3 *SD* from the mean reaction time for each subject) and error trials, which were rare and did not indicate any speed-accuracy tradeoffs, were excluded from the main analyses.

[9] Participant information (age and gender) was missing for two people included in the analysis.

significant, $F(1,51) = 0.001$, $p = .99$, $\eta_p^2 = .000$. The means and standard errors are plotted in Fig. 3.

Pairwise post hoc *t*-tests confirmed that the reaction time comparison between false beliefs ($M = 2111$ ms, $SD = 571$ ms) and false notes ($M = 2243$ ms, $SD = 537$ ms) was significant, $t(51) = 3.05$, $p = .004$, $d = 0.423$, as was the difference between true beliefs ($M = 1991$ ms, $SD = 571$ ms) and true notes ($M = 2124$ ms, $SD = 529$ ms), $t(51) = 2.23$, $p = .030$, $d = 0.309$. There were also differences between false and true notes, $t(51) = 2.43$, $p = .019$, $d = 0.337$, and false and true beliefs, $t(51) = 3.57$, $p < .001$, $d = 0.495$.

The same advantage for mental states seen in Experiment 1 was observed again in a format-stressed version of the hybrid-task. The difference in percent correct between beliefs and notes appeared larger under "false" than "true" conditions, consistent with the pattern seen in Experiment 1 and suggesting that accuracies in the "true" conditions may have been approaching a performance ceiling, although, unlike Experiment 1, the reaction time advantages for belief were equally large under both truth-value conditions. The evidence from this experiment suggests that even when representational properties of notes are stressed, encouraging deployment of representational knowledge, the representational processing theory is not supported. In contrast, the results appear to provide evidence for specialization of mental state processing.

Is it possible that participants were slower for notes because they had to compute belief content in order to determine note content?[10] That is, given that the content of the note is not visible to participants, and that the agent is the one to write the note, an alternative explanation for Experiments 1 and 2 is that participants inferred the mental state of the agent and then copied the content of the mental state over into their representation of the note (either online when they see the agent create the note or offline when they are prompted by the probe). Pre-empirically, this possibility seems to prioritize belief reasoning (after all, beliefs are not visible either) and to assume some degree of specialized computation, treating belief ascription as the basis for inferring note content (see Section 6.4. for a related idea). Additionally, nothing about the general structure of the task requires participants to compute belief in order to compute the content of the note. In a study investigating the ability to reason about false drawings in children with autism, children watched an experimenter draw a picture of an object which was then replaced by a different object. Participants were able to reason about the content of the false drawing despite being impaired on a matched false-belief task, suggesting that their ability to reason about the drawings did not depend on first computing the experimenter's belief (Charman & Baron-Cohen, 1992). Similarly, there are likely non-mentalistic routes to computing the content of the note in the current task in addition to using mental states.

But given this is an empirical issue, we address this counterhypothesis in two remaining studies. Participants were given new instructions with a similar cover story but were now told that after the woman hides her friend's purse, she (a) takes out a piece of paper and writes in a note where in the room the purse is located and (b) calls back her friend and says on the phone where in the room the purse is located. Critically, instead of having to reason about the belief of the woman hiding the purse (the task in the belief condition of the first two experiments), participants had to reason about the belief of the friend returning to search for the purse. If participants were using the woman's belief to infer the content of the note in the first two experiments, requiring extra processing, that cost should now be equal since the contents of the friend's belief are also derived from the woman's belief. We reran Experiments 1 and 2 to test this alternative possibility.

4. Experiment 3

4.1. Method

4.1.1. Participants

We tested 72 participants, all undergraduates, participating for course credit. Of the 72 participants, 9 failed to meet accuracy criteria and were not analyzed, leaving the final sample at $N = 63$ (45 females and 18 males; $M_{age} = 21.95$ years, $SD_{age} = 4.24$ years).[11]

4.1.2. Design and procedure

The design and procedure were nearly identical to the previous experiments (see Appendices A and C for complete instructions and text of subtitles). However, the subtitles presented with the initial practice video were modified so that instead of instructions to reason about the actress's belief and the note as in Experiment 1 and 2, participants now had to reason about the friend's belief and the note. Critically, the antecedents to establishing the friend's belief and the note's content in the video stimuli and instructions were matched. After practice but before the test phase, participants were told that when reading the probes they should interpret "the friend" to mean "the friend Bridget calls" and "the note" to mean "the note Bridget writes," ("Bridget" being the name of the protagonist). As in Experiment 1, participants received belief or note test probes (e.g., "The friend thinks the purse is in the right drawer"; "The note shows the purse is in the right drawer") or one of several kinds of filler probes (e.g., "It is true that the purse has switched drawers"; "It is true that the purse is in the left drawer"; "It is true that Bridget wrote a note"; "It is true that Bridget called her friend").

4.2. Results and discussion

Percent correct was entered into a two-way repeated-measures ANOVA, revealing a main effect of representation type, $F(1,62) = 4.89$, $p = .031$, $\eta_p^2 = .073$, such that percent correct was significantly higher for beliefs than notes, and a main effect of truth value, $F(1,62) = 18.7$, $p < .001$, $\eta_p^2 = .232$, such that percent correct was significantly

[10] We thank Leda Cosmides and two anonymous reviewers for this suggestion.

[11] Participant information (age and gender) was missing for two people that were included in the analysis.

higher for true than false representations. These main effects were qualified by a representation-type by truth-value interaction, $F(1,62) = 10.8$, $p = .002$, $\eta_p^2 = .148$. The means and standard errors are plotted in Fig. 2.

Pairwise post hoc *t*-tests confirmed that the accuracy comparison between false beliefs ($M = 94.3\%$, $SD = 0.10$) and false notes ($M = 89.8\%$, $SD = 0.12$) was significant, $t(62) = 3.20$, $p = .002$, $d = 0.432$, although there was no difference between true beliefs ($M = 95.7\%$, $SD = 0.07$) and true notes ($M = 96.7\%$, $SD = 0.06$), $t(62) = 1.31$, $p = .196$, $d = 0.235$. There was also a difference between false and true notes, $t(62) = 4.97$, $p < .001$, $d = 0.732$, but no difference between false and true beliefs, $t(62) = 1.16$, $p = .250$, $d = 0.164$.

Reaction times were subjected to a two-factor repeated-measures ANOVA, revealing a main effect of representation type, $F(1,62) = 5.15$, $p = .027$, $\eta_p^2 = .077$, such that reaction times to beliefs were significantly faster than reaction time to notes, and a main effect of truth value, $F(1,62) = 52.8$, $p < .001$, $\eta_p^2 = .460$, such that reaction times were significantly faster to true than false representations. The representation-type by truth-value interaction was not significant, $F(1,62) = 0.279$, $p = .599$, $\eta_p^2 = .004$. The means and standard errors are plotted in Fig. 3.

Pairwise post hoc *t*-tests confirmed that the reaction time comparison between false beliefs ($M = 2355$ ms, $SD = 696$ ms) and false notes ($M = 2435$ ms, $SD = 709$ ms) was significant, $t(62) = 2.00$, $p = .050$, $d = 0.189$. No difference emerged between true beliefs ($M = 2139$ ms, $SD = 619$ ms) and true notes ($M = 2190$ ms, $SD = 662$ ms), $t(62) = 1.34$, $p = .185$, $d = 0.174$, although there were differences between false and true notes, $t(62) = 6.06$, $p < .001$, $d = 0.762$, and false and true beliefs, $t(62) = 5.13$, $p < .001$, $d = 0.688$.

The results failed to support the alternative hypothesis that an advantage for processing beliefs compared to notes was an artifact of unequal task demands, and provided added support for the initial hypothesis of specialized belief processing. To further increase confidence that the alternative account could not explain the initial findings, we re-ran Experiment 2 using a modified format-stressed task that referred to the friend's belief.

5. Experiment 4

5.1. Method

5.1.1. Participants

Seventy participants, all undergraduates participating for course credit, were tested. Of the 70 participants, seven failed to meet accuracy criteria and their data were not analyzed, leaving the final sample at $N = 63$ (46 females and 17 males; $M_{age} = 22.9$ years, $SD_{age} = 6.49$ years).

5.1.2. Design and procedure

The design and procedure were nearly identical to previous experiments. As in Experiment 2, the belief and note probes started, "In her mind, the purse…" and "In the note, the purse…," respectively, in order to highlight format. As in Experiment 3, participants were asked to reason about the friend's belief and the note's content.

5.2. Results and discussion

Percent correct was entered into a two-way repeated-measures ANOVA, revealing a main effect of representation type, $F(1,62) = 5.77$, $p = .019$, $\eta_p^2 = .085$, such that percent correct was significantly higher for beliefs than notes, and a main effect of truth value, $F(1,62) = 22.0$, $p < .001$, $\eta_p^2 = .262$, such that percent correct was significantly higher for true than false representations. The representation-type by truth-value interaction was not significant, $F(1,62) = 2.72$, $p = .104$, $\eta_p^2 = .042$. The means and standard errors are plotted in Fig. 2.

Pairwise post hoc *t*-tests confirmed that the accuracy comparison between false beliefs ($M = 90.1\%$, $SD = 0.13$) and false notes ($M = 83.8\%$, $SD = 0.23$) was significant, $t(62) = 2.20$, $p = .031$, $d = 0.278$, although there was no difference between true beliefs ($M = 96.1\%$, $SD = 0.06$) and true notes ($M = 94.7\%$, $SD = 0.09$), $t(62) = 1.23$, $p = .224$, $d = 0.155$. There were, however, differences between false and true notes, $t(62) = 3.74$, $p < .001$, $d = 0.471$, and false and true beliefs, $t(62) = 3.90$, $p < .001$, $d = 0.491$.

Reaction times were subjected to a two-factor repeated-measures ANOVA, revealing a main effect of representation type, $F(1,62) = 6.70$, $p = .012$, $\eta_p^2 = .097$, such that reaction times to belief probes were significantly faster than reaction times to note probes, and a main effect of truth value, $F(1,62) = 53.2$, $p < .001$, $\eta_p^2 = .462$, such that reaction times were significantly faster to true than false representations. These main effects were qualified by a representation-type by truth-value interaction, $F(1,62) = 8.07$, $p = .006$, $\eta_p^2 = .115$. The means and standard errors are plotted in Fig. 3.

Pairwise post hoc *t*-tests confirmed that the reaction time comparison between false beliefs ($M = 2417$ ms, $SD = 592$ ms) and false notes ($M = 2584$ ms, $SD = 643$ ms) was significant, $t(62) = 3.29$, $p = .002$, $d = 0.414$. However, no difference emerged between true beliefs ($M = 2223$ ms, $SD = 616$ ms) and true notes ($M = 2235$ ms, $SD = 520$ ms), $t(62) = 0.33$, $p = .741$, $d = 0.042$, while there were differences between false and true notes, $t(62) = 7.73$, $p < .001$, $d = 0.974$, and false and true beliefs, $t(62) = 4.10$, $p < .001$, $d = 0.516$.

These results provide further evidence that advantages to process beliefs compared to notes in Experiments 1 and 2 were due to specialized belief processing, not unequal task demands in reasoning about beliefs and notes.

6. General discussion

The results of Experiments 1–4 showed a consistent advantage for processing the content of a belief over the content of a note. Even when representational properties were made explicit (Experiments 2 and 4) the faster reaction time and higher accuracy effects for belief probes compared to note probes persisted. These results demonstrate for the first time a dissociation in processing of mental and linguistic representations, supporting the prediction derived from the mental state processing theory that reaction time and accuracy advantages emerge for belief probes because of a specialized processing mechanism for theory of mind, but they failed to support broader specialization theories proposing a capacity to metarepresent or to specifically represent propositional representations.

Compared to previous studies, the current set of experiments introduced a number of changes that establish a clearer and stronger test of the two theories under consideration. First, the experiments addressed the falseness problem by (a) substituting in a linguistic representation that would be true or false under the same conditions as a belief and (b) including "true" representation conditions. Second, the tests handled the association problem by adopting a dissociation approach and using a task modeled on the false-photo task that tightly matched conditions for reasoning about notes and beliefs. Third, these studies addressed the content problem by comparing mental and linguistic representations, which share a deep representational similarity: both express propositions. This representational closeness strengthened the test considerably compared to prior tests using pictorial representations. By speaking to the triplet of problems, the new tasks have the advantage of addressing key criticisms from researchers on both sides of the theoretical debate, producing a task that we believe significantly improves upon those used in previous studies. The current studies also incorporated a more sensitive reaction time method compared to less discriminating pass/fail measures that are commonly used. Taken together, these new features establish a stronger, more sensitive test of the broad and narrow specialization theories, with the findings reported here providing evidence for special-purpose machinery for theory of mind.

6.1. Potential objections considered

6.1.1. Is there still a falseness problem?

Are notes conceptually problematic like photos? Could there be inherit differences in falsity conditions between mental and linguistic representations? First, notice that the falseness problem predicts that false photos should be *easier* to process than mental states; false belief should require reconciling a false representation with the current state of affairs that it conflicts with, whereas photos suffer no conflict because they refer to a true past state of affairs, not to the present (Perner & Leekam, 2008). Our results with notes show that they are harder to process than beliefs, suggesting this conceptual difference is not in play. Besides this empirical point, on conceptual grounds, it is not obvious that notes differ from beliefs on falseness; like beliefs, they can be genuinely false because they refer to the current state of reality. Therefore, the criticism of photos does not appear to apply to notes.

6.1.2. Other potential objections

Unlike beliefs, which are abstract, public representations like notes are available to the senses. From this consideration, we ought to predict that notes should be easier, not harder than calculating belief. Given that beliefs were processed more efficiently than notes, the criticism from abstractness seems unable to account for the results.

Another potential issue is that although participants see the person reach for the paper and write the note, they never get to see its actual content, and it could be argued that this might contribute to slower reaction times and lower accuracy. However, this is an important feature of the task because it controls for the fact that people never get to see the contents of a person's mental state. Because the contents of mental states are unobservable and have to be inferred, the same criterion has to hold in the note condition.

Are there differences in how the instructions direct attention to notes compared to beliefs? The instructions drew attention to the note by explicitly saying that she "leaves a note for her friend indicating where in the room the purse is located," but do not equivalently direct attention to belief by saying "she thinks about where in the room the purse is located." We suspect this difference benefited reasoning about notes, and suggest the current results may underestimate processing advantages for belief.

Given that belief probes in Experiments 1 and 3 contained the verb "thinks," might participants have used "thinks" contrastively with "knows," leading them to interpret "thinks" to mean the agent had a false belief?[12] If so, this might speed up reaction times or increase accuracy on false-belief trials, creating a processing advantage compared to false notes, and slow reaction times or decrease accuracy on true-belief trials, eliminating a processing advantage compared to true notes. Critically, such a representation-type by truth-value interaction on accuracy and reaction time would only be predicted in Experiments 1 and 3, in which "thinks" was used in the belief probes (Experiments 2 and 4 used "In her mind…"). In fact, of the eight possible interactions (four experiments each with two analyses, one for accuracy and one for reaction time), six were observed, with at least one interaction in each of the four experiments. Of the two interaction analyses that clearly failed to reach significance, one was for reaction time in Experiment 2 and the other for reaction time in Experiment 3. Although we suspect that the mind does interpret "thinks" to imply an agent may be mistaken or unsure under some conditions, the tasks were not designed to detect this "contrastive" effect and no clear pattern emerged supporting this alternative account. We conclude that "contrastive" effects are not the source of the main differences here, which appear to be better explained by the "specificity" effects the task was designed to test.

Taken together, comparing notes to belief has a number of virtues: they are generally well matched, and where differences exist, the consequences of these differences generally predict no advantage or an advantage opposite from the observed advantages for belief, suggesting the experiments reported here may underestimate the "specificity" effects.

[12] We thank an anonymous reviewer for suggesting this possibility.

6.2. Theory of mind, metalinguistic knowledge, and correlations

The current studies are not the first to explore a relationship between mental and linguistic representations.[13] Researchers advocating a representational processing approach argue that theory of mind and metalinguistic knowledge are conceptually related because both depend on understanding the distinction between *what is represented* and *how it is represented* (Doherty, 2000; Doherty & Perner, 1998). For example, in the false-belief task, the child has to understand that although Sally has a belief about the ball (what is represented: the ball), she represents it as being in a location (how it is represented: "in the basket") that is different from the child's own belief (how it is represented: "in the box"). Likewise, understanding linguistic representations such as synonyms involves recognizing that two words can have the same meaning despite having different surface forms. For instance, "rabbit" refers to the animal with big ears, but "bunny" also refers to the animal with big ears. That is, one has to appreciate that what is represented can be the same (the animal with the big ears) even though how the target is represented can be different ("rabbit" or "bunny"). A similar analysis can be run on homonyms, except that form remains constant (e.g., "bat") and meaning varies (e.g., the flying mammal or baseball equipment).

If understanding synonyms, homonyms, and belief all depend on the ability to understand this aspect of representation, then they should develop along a similar developmental timetable, leading to the prediction that performance on tasks of metalinguistic knowledge (synonyms and homonyms) should correlate with performance on tasks assessing false-belief knowledge, and there is indeed evidence for such a relationship (Doherty, 2000; Doherty & Perner, 1998; Garnham et al., 2000; Perner, Stummer, Sprung, & Doherty, 2002). Yet in light of the current findings, it is less clear that the correlations between performance on the false-belief task and the synonym and homonym tasks in 3- to 5-year-olds are driven by a common ability to understand representation as originally suggested (Doherty, 2000; Doherty & Perner, 1998; Perner et al., 2002). Given the processing dissociations reported in Experiments 1–4, is there a different explanation for the associations reported from the development literature?

The current findings with adults and the prior results with preschoolers are consistent with at least two possibilities. First, in younger children, belief reasoning and metalinguistic knowledge may be related because of shared executive demands in the false-belief and synonym/homonym tasks, which act as limiting factors on performance, with changes in performance yoked as executive function systems develop (Garnham et al., 2000). Second, it is possible that a more broadly specialized system in early development gives way to two separate mechanisms at some point between preschool and early adolescence. We flesh out the executive function explanation next, and return to changes in specialization in Section 6.3.

One candidate executive demand that factors in both the false-belief task (Carlson, Moses, & Breton, 2002; Carlson, Moses, & Hix, 1998; Leslie & Polizzi, 1998) and the synonym/homonym tasks (Garnham et al., 2000) is the need to inhibit salient information. In the false-belief task, children must inhibit a representation about the actual location of the object in order to predict that someone will search in an empty location that they falsely believe contains the object. In the synonym task, children must inhibit the tendency to repeat the most recently named word in the synonym pair and state the other word, while in the homonym tasks, they must inhibit pointing to the picture most recently pointed to (one member of the homonym pair) and instead point to the picture that picks out the other homonym. If inhibition is required in these tasks, then performance should steadily improve and positively correlate over the preschool years as inhibitory control develops. The current results, which support two independent systems for theory of mind and metalinguistic knowledge, more closely favor a 3rd variable account that appeals to executive function over an explanation that appeals to a common capacity for understanding representation.

Further evidence for the executive function explanation comes from studies of bilinguals. Learning that different word forms *across* languages can have the same meaning poses similar challenges to learning that different word forms *within* a language can have the same meaning. Research with bilinguals consistently shows that they outperform their monolingual peers on executive functioning tasks (Bialystok & Craik, 2010; Kovács & Mehler, 2009) and that this executive function advantage is related to their advanced development on false-belief tasks (Bialystok & Senman, 2004; Goetz, 2003; Kovács, 2009). These studies strengthen the case that executive function plays a supporting role in reasoning about both linguistic and mental representation, providing further evidence for the executive function explanation.

6.3. Early- or late-developing specialization?

The current results suggest that by early adulthood, specializations for processing belief are in place; however, they do not illuminate whether specialization extends back into early development, and if it does, how far back. Because we looked at adults and not preschoolers or younger children, an open question is whether the mechanisms are specialized in virtue of expertise (accumulated experience over development) or due to early and reliably developing specialization that forms the foundation of development rather than its outcome. To be clear, this is not an argument *for* narrow specialization; at issue is *when* and *how* narrow specialization develops.

[13] There is extensive work on the relationship between language and theory of mind, including philosophical claims about belief concepts and natural language (Davidson, 1975; Sellars, 1956) and psychological research linking (a) complementation syntax to theory of mind development (de Villiers & Pyers, 2002) and (b) theory of mind to language learning and communication (Bloom, 1997; Grice, 1989; Sperber, 2000). Others place emphasis on the independence of language and theory of mind (Varley & Siegal, 2000). Though important, these debates are beyond the scope of this paper; both the narrow and broad specialization positions are *conceptual accounts* of theory of mind – neither one appeals to linguistic-specific achievements being necessary for the *emergence* of theory of mind, although language may play a role in its *expression*.

An argument in favor of specialized mechanisms that develop early and reliably rests on the data from samples with autism mentioned earlier. If specialization is late developing in virtue of expertise, we might expect differences between children with and without autism to emerge only later in development, when the mechanisms have become specialized in typically developing children. The fact that the gap in theory of mind ability is so pronounced in preschool-age children suggests that what, at least partly, explains the early social maturity of one group and profound social difficulties of the other is the difference in integrity of specialized, early-developing learning mechanisms.

In addition to arguments for early specialization, there are other arguments that can be raised against late specialization. For example, for an expertise system to accumulate more experience with mental than non-mental representations, it would have to be sensitive to differences between one type of representation (beliefs) and others (notes, maps, signs). If it is sensitive to any differences, then there must be a system narrowly specialized to at least make those distinctions in the first place. What happens if an expertise system does not make distinctions between types of representations? Critically, training and any benefits accrued with one type of representation should transfer to other types of representations. But if there is transfer, then it seems difficult to acquire more experience with one kind of representation than another because practice with any one representation should get folded into experience with representations *in general*. Whether and under what conditions this argument can be extended to other cases of specializations beyond theory of mind is an open question.[14]

If broad specialization is taken to be distinct from expertise accounts, might the belief reasoning system shift from broad to narrow specialization across development? For the same reasons just outlined, transfer is also expected on the representational processing theory (Iao et al., 2011), which is why broad specialization is also unlikely to be the basis for late, narrow specialization. Given these considerations, it is not clear that general expertise or broadly specialized mechanisms for representations could generate narrow specialization for belief without also generating specialization for all other types of representations. While we think these considerations provide some theoretical support for early specialization over late specialization, empirical evidence will also be needed. Studies with preschoolers comparing false beliefs to false notes or to false signs will be especially informative because they can provide direct evidence for or against early specialization.

We also emphasize that rather than seeing specialization and generalization as opposite in all respects, there is much to be said in favor of the idea that with increased specialization comes some ability to generalize, at least in two respects. First, given that belief reasoning involves concepts like AGENT and BELIEF, the system is expected to generalize across an unbounded set of events involving agents. Second, we suggest there may be some important connections with an idea put forward in the next section (Section 6.4) in which specialized mechanisms designed to process a set of inputs can, as a by-product, process related types of input. While it remains an open question what the combination of specialization and generalization in adults might say about the mechanisms at an earlier point in development, one idea we are sympathetic to is that specialized mechanisms get learning off the ground but development might be better thought of as extending these specialized processes in new ways, either by operating over new types of input or by creating interfaces with other systems that transform representations and regulate behavior in new ways.

6.4. Extending the narrow specialization theory: a by-product account

Are narrow and broad specialization accounts competing or complementary theories? The version of narrow specialization considered here is at odds with the broad specialization account over the cause of associations and the existence of dissociations. These competing explanations of data reflect deeper differences in theory about cognitive architecture and development: Narrow specialization assumes the development of distinct systems for processing mental and non-mental representations, a view denied by broad specialization. We propose it may be possible to extend the narrow specialization theory in a way that more fully explains patterns of associations and dissociations with mental and non-mental representations. On this extended and more complementary proposal, which we call the by-product theory, a specialized mechanism for processing mental states may be *capable of*, though not *designed for*, processing other types of representations. As theorists have pointed out, non-mental representations are similar in many respects to mental representations (see Section 1.1.), and on the by-product account, the theory of mind system may be triggered by non-mental representations.

Differences in accuracies and reaction times like those observed in Experiments 1–4 could emerge when a cognitive system, programmed to process belief and other

[14] An anonymous reviewer noted as a counter-example that one's ability to play chess may be somewhat compromised when playing with unconventional-looking chess sets (see D'Andrade, 1981 for anecdotal confirmation), and based on this, suggested we would not conclude that there are pre-existing narrowly specialized mechanisms for the ability to play with conventional- (vs. unconventional-) looking chess pieces. In the case of chess, it seems likely that early experience with unconventional-looking chess pieces would lead to the opposite outcome – namely, that one's chess-playing ability would then be compromised when playing with conventional-looking pieces. We would argue this is not analogous to the case of mental vs. non-mental representations because there does not appear to be any evidence that adults are more efficient at processing non-mental representations by virtue of greater experience with non-mental than mental representations. Typical participants are exposed to symbols, signs, notes, maps, photos, and other non-mental representations throughout their lives and more so in modern developed society than at any other time and place in history, and although they have never directly perceived a belief, our research suggests that they are still able to process mental representations more efficiently than non-mental representations. The only instance of greater processing efficiency for non-mental than mental representations comes from studies of people with autism. It is far from clear that people with autism encounter environments that provide greater exposure to non-mental than mental representations and, at the same time, that typical adults encounter environments that provide greater exposure to mental than non-mental representations.

mental states, is confronted with input that it was not designed to process, leading to inefficiencies and processing costs (Cohen & German, 2010). At the same time, the by-product theory can account for correlations between mental and non-mental representations given that the same underlying mechanism is activated across representation types. Whereas the standard version of the narrow mental state specialization account predicts dissociations but not associations after performance demands have been controlled for, and the broad representational processing account predicts associations but not dissociations, only the by-product account predicts associations *and* dissociations.

We consider two hypotheses for why inefficiencies arise when specialized mental state processing operates over non-mental representations: the "format-blind" hypothesis and the "social cue" hypothesis. On the "format-blind" hypothesis, narrowly specialized mechanisms may be designed to process representational content, but given that mental states are unobservable, they pay no attention to the medium of the representation. When such mechanisms are confronted with non-mental representations, which have public and concrete representational mediums in addition to their content, this may pose a problem for the system. Registering format may even be required to understand that one is observing a pictorial representation. As Ittelson (1996) observes, failing to appreciate the medium of a pictorial/iconic representation that resembles its referent amounts to mistaking the representation for the real thing.[15] Although this analysis would not apply to linguistic representations, which not only fail to resemble their referents but stand in an arbitrary relationship to them, processing linguistic format may serve other purposes (e.g., individuating representations with the same content but different format, such as directions conveyed verbally in a note or pictorially in a map). On the "format-blind" hypothesis, because specialized mental state processing mechanisms are not designed to handle the format information associated with public representations, they require either (a) other processes to strip the non-mental representation of format information or (b) extra time for the specialized mechanisms to recognize the content of the non-mental representation due to interference from information about medium. Relatedly, rather than being thought of in terms of processing inefficiencies, the format-blind hypothesis could be viewed in terms of a mechanism specialized to process both mental and non-mental representations by entering into different processing modes. On this account, mental representations might turn on a "format-blind" mode while non-mental representations turn on a second mode that strips format in addition to the computations executed in the other mode. On this version of the format-blind hypothesis, longer reaction times would not reflect a loss in efficiency or lack of specialization but rather a different processing mode with extra processing steps.[16]

The "social cue" hypothesis proposes that specialized mechanisms for processing mental states are most strongly activated by the presence of certain social cues, including eye gaze (Baron-Cohen, 1995), morphology and contingent motion (Johnson, 2003), self-propelled motion (Premack, 1990), linguistic cues (Roth & Leslie, 1991), and action toward objects (Wertz & German, 2007, 2013). Non-mental representations do not emit such cues; therefore, whatever non-social input they provide may only weakly activate the specialized mechanisms, and the greater the mismatch between the cues that are available and the cues that are expected by the system, the larger the costs incurred to recognize the cues and reach threshold for activating the system.

6.5. Summary

The purpose of the current investigation was to test whether the scope of the processing domain for mechanisms underlying theory of mind corresponded to the broader domain of representation (the representational processing theory) or the narrower domain of *mental* representation (the mental state processing theory). Even though both mental and linguistic representations express propositions, making them representationally more similar to each other than the mental and pictorial representations used in previous research (e.g., Cohen & German, 2010; Iao & Leekam, 2014), the current investigation revealed a dissociation in which processing was quicker and often more accurate for beliefs than for notes. This pattern persisted even when representational format was emphasized, which should have favored activating representation concepts. These findings suggest that mental representations are special because they are mental, not because they are representations.

Building on the mental state processing theory, which proposes that domain specificity amounts to *restricted* processing of mental states, the by-product theory relaxes this restriction and replaces it with *privileged* processing of mental states along with the capacity to extend processing to non-mental representations. An important direction for future work will be testing for further empirical support of the by-product theory, which handles the existing evidence better than previous accounts.

Appendix A

Instructions prior to practice round of Experiments 1–4:

- "You will see a series of short videos. At the end of each video (when a blue frame appears), you will have to point to the location where you think an object is. The experimenter will tell you which object to track."
- Experimenter verbally tells participants that the object they will be tracking is a woman's purse.

[15] Even pictorial representations that do not resemble their referents likely require encoding of the medium to prevent misrepresentation (of a misrepresentation). For example, if one draws a popsicle, but it looks like a balloon, failing to notice the medium should cause one to mistake the representation for a real balloon rather than a real popsicle. Ittelson's point, then, seems to apply to not only true/accurate pictorial representations but also false/inaccurate ones, such as those commonly used in theory of mind studies.

[16] We thank an anonymous reviewer for this suggestion.

- "At some point during each video a statement will appear on the screen, and you will have to respond 'yes' or 'no.' You will press the 'y' key if you think the answer to the statement is 'yes' and the 'n' key if you think that the answer is 'no.' Please answer as quickly and accurately as possible while avoiding errors."
- "Each new video is completely separate and independent from the previous ones. At all times you should interpret 'left' and 'right' from your point of view."
- "The first three videos are practice. You will receive feedback to your answers ONLY during the three practice videos. The first practice video includes a background story that applies to all of the videos you will see. You will only get the background story on the FIRST practice video, so please pay careful attention. Also, the first video will be played at normal speed. All videos that follow will be played at a faster speed."

Instructions after practice round of Experiments 3 and 4 ONLY:

- "Throughout the task, 'The friend' = the friend Bridget calls; 'The note' = the note Bridget writes."

Appendix B

Subtitles overlaid on first practice video of Experiments 1 and 2:

- At the start of the video, before frame 1 of Fig. 1: "Bridget is a reporter for a school newspaper. Her friend, who stopped by earlier today, forgot her purse. Bridget is waiting for her to come pick it up. While she's waiting, she gets a phone call."
- After the start of the video but still before frame 1 of Fig. 1: "Bridget has to leave, so she puts the purse away and leaves a note for her friend indicating where in the room the purse is located."
- Between frames 2 and 3 of Fig. 1: "After Bridget leaves, her officemate Barry stops by to drop off a book."

Appendix C

Subtitles overlaid on first practice video of Experiments 3 and 4:

- At the start of the video, before frame 1 of Fig. 1: "Bridget is at work and has to leave soon. She is holding onto a friend's purse and waiting for the friend to come pick it up. While she's waiting, she gets a phone call from her friend."
- After the start of the video but still before frame 1 of Fig. 1: "Bridget has to leave now, so she hides the purse. To make sure her friend finds it, she (a) takes out a piece of paper and writes in a note where in the room the purse is located and (b) calls back her friend and says on the phone where in the room the purse is located."
- Between frames 2 and 3 of Fig. 1: "After Bridget leaves, her officemate Barry stops by to drop off a book."

References

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention. *Journal of Cognitive Neuroscience, 21*(6), 1179–1192.

Avis, J., & Harris, P. L. (1991). Belief–desire reasoning among Baka children: Evidence for a universal conception of mind. *Child Development, 62*, 460–467.

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37–46.

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., et al. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Sciences, 280*(1755), 20122654.

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science, 19*, 19–23.

Bialystok, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*, 562–579.

Bloom, P. (1997). Intentionality and word learning. *Trends in Cognitive Sciences, 1*, 9–12.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*(1), B25–B31.

Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the mind or predicting signal-dependent action? Performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development, 6*(2), 259–283.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*, 187–192.

Callaghan, T. C., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., et al. (2005). Synchrony in the onset of mental-state reasoning. *Psychological Science, 16*, 378–384.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.

Charman, T., & Baron-Cohen, S. (1992). Understanding drawings and beliefs: A further test of the metarepresentation theory of autism: A research note. *Journal of Child Psychology and Psychiatry, 33*(6), 1105–1112.

Charman, T., & Baron-Cohen, S. (1995). Understanding photos, models, and beliefs: A test of the modularity thesis of theory of mind. *Cognitive Development, 10*(2), 287–298.

Cohen, A. S., & German, T. C. (2010). Reaction time advantages for calculating beliefs over public representations signal domain specificity for 'theory of mind'. *Cognition, 115*, 417–425.

Corballis, M. (2003). Recursion as the key to the human mind. In K. Sterelny & J. Fitness (Eds.), *From mating to mentality: Evaluating evolutionary psychology* (pp. 155–171). New York: Psychology Press.

Crane, T. (2009). Is perception a propositional attitude? *Philosophical Quarterly, 59*, 452–469.

D'Andrade, R. G. (1981). The cultural part of cognition. *Cognitive Science, 5*, 179–195.

Davidson, D. (1975). Thought and talk. In S. Guttenplan (Ed.), *Mind and language*. Oxford: Oxford University Press.

de Villiers, J. G., & Pyers, J. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*, 1037–1060.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language, 27*, 367–392.

Doherty, M. J., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development, 13*, 279–305.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford: Oxford University Press.

Frege, G. (1892/1980). On sense and reference. In P. Geach, & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Oxford: Blackwell.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Psychology Press.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 358*, 459–473.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*, 531–534.

Garnham, W. A., Brooks, J., Garnham, A., & Ostenfeld, A. M. (2000). From synonyms to homonyms: Exploring the role of metarepresentation in language understanding. *Developmental Science, 4*, 428–441.

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6*, 1–15.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.

Goodman, N. (1976). *Languages of art*. Indianapolis, IN: Hachett Publishing Co.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Harris, P. L. (2006). Social cognition. In W. Damon, R. Lerner, D. Kuhn, & R. Siegler (Eds.). *Handbook of child psychology: Cognition, perception and language* (6th ed., Vol. 2, pp. 811–858). New York: John Wiley.

Iao, L. S., & Leekam, S. R. (2014). Nonspecificity and theory of mind: New evidence from a nonverbal false-sign task and children with autism spectrum disorders. *Journal of Experimental Child Psychology, 122*, 1–20.

Iao, L., Leekam, S., Perner, J., & McConachie, H. (2011). Further evidence for nonspecificity of theory of mind in preschoolers: Training and transferability in the understanding of false beliefs and false signs. *Journal of Cognition and Development, 12*, 56–79.

Ittelson, W. H. (1996). Visual perception of markings. *Psychonomic Bulletin & Review, 3*, 171–187.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society, 358*, 549–559.

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science, 12*, 48–54.

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, 106*, 6556–6560.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834.

Leekam, S., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40*, 203–218.

Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal of Developmental Psychology, 26*, 485–497.

Leslie, A. M., Friedman, O., & German, T. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*, 528–533.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology, 50*, 45–85.

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). New York: Cambridge University Press.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1*, 247–253.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology, 44*, 523–531.

Lopes, D. (1996). *Understanding pictures*. Oxford: Oxford University Press.

Lopes, D. M. (2003). Pictures and the representational mind. *Monist, 86*, 35–52.

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

Perini, L. (2005). The truth in pictures. *Philosophy of Science, 72*, 262–285.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition, 57*, 241–269.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience, 1*(3–4), 245–258.

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false. *Quarterly Journal of Experimental Psychology, 61*, 70–89.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125–137.

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science, 333*, 474–477.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development, 17*, 1451–1472.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1–16.

Roth, D., & Leslie, A. M. (1991). The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Developmental Psychology, 9*, 315–330.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development, 77*, 1034–1049.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*, 87–124.

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

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & Language, 14*, 131–153.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172–1196.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.). *Minnesota studies in philosophy of science* (Vol. 1). Minneapolis: University of Minnesota Press.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (1st ed., pp. 117–138). Oxford: Oxford University Press.

Stone, V. E., & Gerrans, P. (2006). What's domain-specific about theory of mind? *Social Neuroscience, 1*, 309–319.

Suddendorf, T. (1999). The rise of the metamind. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 218–260). London: Oxford University Press.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychological Science, 18*, 580–586.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*, 552–564.

Varley, R., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Current Biology, 10*, 723–726.

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

Wertz, A. E., & German, T. C. (2007). Belief-desire reasoning in the explanation of behavior: Do actions speak louder than words? *Cognition, 105*, 184–194.

Wertz, A. E., & German, T. C. (2013). Theory of mind in the wild: Toward tackling the challenges of everyday mental state reasoning. *PLoS ONE, 8*(9), e72835.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition, 35*, 41–68.