Cognition 115 (2010) 417–425

Contents lists available at ScienceDirect

Cognition

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

A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'

Adam S. Cohen, Tamsin C. German *

Department of Psychology, University of California, Santa Barbara, United States

ARTICLE INFO

Article history:
Received 19 April 2009
Revised 5 February 2010
Accepted 1 March 2010

Keywords:
Theory of mind
Domain specificity

ABSTRACT

In a task where participants' overt task was to track the location of an object across a sequence of events, reaction times to unpredictable probes requiring an inference about a social agent's beliefs about the location of that object were obtained. Reaction times to false belief situations were faster than responses about the (false) contents of a map showing the location of the object (Experiment 1) and about the (false) direction of an arrow signaling the location of the object (Experiment 2). These results are consistent with developmental, neuro-imaging and neuropsychological evidence that there exist domain specific mechanisms within human cognition for encoding and reasoning about mental states. Specialization of these mechanisms may arise from either core cognitive architecture or via the accumulation of expertise in the social domain.

1. Introduction

A question of key importance for cognitive science concerns the extent to which interpretations and inferences about social agents' behavior are yielded by cognitive mechanisms that are specialized for that purpose (hereafter, 'domain specific' mechanisms), rather than by more general inferential machinery. While some theories propose such domain specific mechanisms as part of the core cognitive architecture for belief-desire reasoning (e.g. Leslie, Friedman, & German, 2004), others propose more general 'executive' inference processes handle mental state content as just one of a broad range of functions covering a variety of content types (e.g. Russell, 1999).

Demonstrations of domain specificity for inferences in the social domain typically rely on comparison of tasks matched closely for general executive demands, but which involve different conceptual content. Investigations of specialization in social attention compare, for example, the effects of centrally presented directional social cues (e.g. eye gaze) with directional non-social cues (e.g. arrows; Ristic, Friesen, & Kingstone, 2002) on otherwise identical cuing paradigms.

An important sub-domain of social cognition is the capacity for belief-desire reasoning, widely studied via the use of tasks assessing the capacity to attribute false beliefs, (Baron-Cohen, Leslie, & Frith, 1985; Wellman, Cross, & Watson, 2001; see Bloom & German, 2000). In this task participants predict the action that follows when an agent's belief goes out of date, and domain specificity of that inference process is diagnosed by comparison to tasks that share the same structural features but that involves no mental content, such as false photographs (Zaitchik, 1990), maps (Leslie & Thaiss, 1992), drawings (Charman & Baron-Cohen, 1992) or signs (Leekam, Perner, Healey, & Sewell, 2008). Divergent performance across the two tasks, given shared general structure, cannot be attributed to the shared general features which might otherwise explain performance in isolation (see e.g. German & Hehman, 2006).

Developmental evidence suggests that preschool children perform similarly on tasks with this general structure (false beliefs, false photos, false maps/signs) although the

* Corresponding author. Address: Department of Psychology, University of California, Santa Barbara, CA 93106-9660, United States. Fax: +1 805 893 4303.
E-mail address: german@psych.ucsb.edu (T.C. German).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.03.001

pattern of inter-correlations suggests stronger association between false beliefs and false signs than between false beliefs and false photos (Perner & Leekham, 2008). The evidence, then, from typical development on these tasks has yet to produce any clear sign of domain specificity for belief processing.

Despite structural parallels and shared general demands of false belief and false photo tasks, which likely accounts for developmental change in this domain (see e.g. Yazdi, German, Defeyter, & Siegal, 2006) there is nonetheless a striking dissociation between the tasks for children with a diagnosis of autism, who fail with false belief content but perform at ceiling with false photos (Leslie & Thaiss, 1992). This evidence has been interpreted as supporting the domain specificity of mechanisms for belief-desire reasoning, although other authors dispute the validity of the false photograph task as an appropriate control for false beliefs (e.g. Perner & Leekham, 2008). Evidence that children with autism appear to fail tasks assessing inferences about false signs (Bowler, Briskman, Gurvidi, & Fornells-Ambrojo, 2005) has been advanced by these authors as an indication that difficulty in both typical and atypically developing populations across these tasks may be the result of the requirement to handle the general concept of 'representation'.[1]

The false belief–false photograph comparison has featured in functional imaging investigations of the domain specificity of theory of mind, principally using fMRI, (Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009). Most prior imaging studies, using a broad range of tasks and materials, showed that a suite of areas are more active in tasks with 'theory of mind' content than in control tasks. These areas include the medial prefrontal cortex (mPFC), the temporal poles and the temporal parietal junction (TPJ; e.g. Fletcher et al., 1995; Gallagher et al., 2000; Gallagher, Jack, Roepstorff, & Frith, 2002; German, Niehaus, Roarty, Giesbrecht, & Miller, 2004; Saxe & Powell, 2006; Saxe & Wexler, 2005; see Amodio and Frith (2006), Frith and Frith (2006) for a review). Of this suite of areas, right temporal parietal junction (rTPJ) appears to show the strongest signal when beliefs and photos (or signs) are compared (Perner et al., 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006), leading to the suggestion that it is rTPJ that implements a domain-specific component of the belief-desire reasoning system.

Neuropsychological evidence supports this general conclusion. While damage to frontal areas is sometimes associated with theory of mind deficits (Rowe, Bullock, Polkey, & Morris, 2001; Stone, Baron-Cohen, & Knight, 1998; Stuss, Gallop, & Alexander, 2001), there are cases where selective damage to mPFC leaves theory of mind inferences intact (Bach, Happé, Fleminger, & Powell, 2000; Bird, Castelli, Malik, Frith, & Husain, 2004). Moreover, patients with frontal lesions fare better on theory of mind tasks designed to reduce language and other processing demands, a manipulation that does not help a small group of patients with left temporo-parietal (lTPJ) lesions, who remain impaired even on these 'lower demand' tasks (Apperly, Samson, Chiavarino, & Humphreys, 2004). In further investigations, lTPJ patients were shown to have problems that extended to reasoning about false photograph tasks (Apperly, Samson, Chiaravino, Bickerton, & Humphreys, 2007). To date, no patients with rTPJ lesions have been tested, and it therefore remains possible that while lTPJ is recruited for both beliefs and other kinds of representation, lesions to rTPJ would impair belief inferences only.

In the current study we investigate the possibility that domain specificity might manifest in belief-desire reasoning in adults, when sensitive measures of performance are used. Accuracy on theory of mind and matched photo tasks are for the most part at ceiling in children older than 5 or 6 years and into adulthood, and thus the current investigation used reaction times to unpredictable probes about beliefs and maps as a more sensitive measure to assess the relative readiness with which the cognitive system makes each kind of inference (see e.g. Apperly, Riggs, Simpson, Samson, & Chiavarino, 2006; Cohen & German, 2009).

We propose that observing different response latencies between belief and map inference tasks would qualify as a signature for domain-specific processing.[2] A domain-general account that posits a common processing mechanism for beliefs and other kinds of representations would sit less easily with such a pattern of responses, if such a processing system is truly blind to the type of content being processed.

In the experiment that follows we present participants with videos of event sequences in which they must track a specific object that appears in the video. We compare participants' responses to probes that appear on some trials asking about the contents of either a belief held by one of the protagonists or a map that one of the protagonists draws during the episode. These 'belief' and 'map' probes were interlaced with a number of other probe types as fillers and were presented randomly so that experimental

[1] The critique of the false photograph task is based on the claim that a photograph that has gone out of dates is not 'false' in the same way as is an outdated belief. While beliefs are 'about' the current situation, an outdated photograph is 'about' the past situation of which it was taken and is a true representation of this past situation (Perner & Leekham, 2008). A false map or sign is a better control for the case of false representation, according to this view, because maps and signs are 'about' the current state of the world, and when they go out of date they are false in the same way as are beliefs; they *mis*represent the current states of affairs. Of course, this is only true because people are *supposed to believe what maps and signs say*, and thus the argument simply introduces the concept of 'belief' into the definition of the concept of representation. If signs rely on the concept of belief in order to achieve *mis*represention, the pattern of correlation in typical development and autism in which there may be a problem 'representing' signs as well as beliefs is rather unsurprising. We are thus unmoved by the argument that photographs are not an appropriate control for beliefs in the domain of autism. Nonetheless, to have the most conservative test of our hypotheses here, we adopt a 'false map' versus 'false belief' comparison in Experiment 1.

[2] Reaction time differences between otherwise closely matched tasks makes for compelling evidence for some specialization in the processing stream, but is not a *necessary* condition for domain-specificity. Equivalent RTs across two tasks might obtain even when processing is driven by different mechanisms. For instance, cuing effects induced by centrally presented eye gaze and arrow cues are not significantly different despite evidence that orienting to gaze and arrows are underpinned by different neural mechanisms (Ristic et al., 2002), Kingstone, Friesen, & Gazzaniga, 2000), and show subtly different behavioral effects under circumstances where cues are counter-predictive (Friesen, Ristic, & Kingstone, 2004).

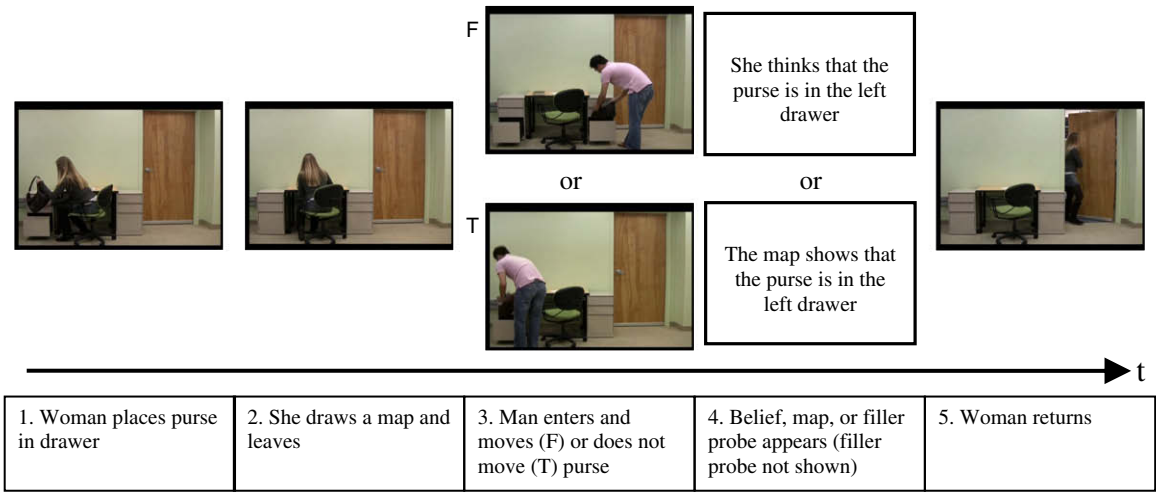

Fig. 1. The sequence of events in the video stimuli. In each film, an actress places the key object in one of the two drawers (1). She then draws a map indicating the location of the object (2). In her absence, a second character enters and during the act of putting a second object in the drawer, either moves the target object (false conditions) or leaves it where it is (true conditions; 3). At this point the video is interrupted with either a belief probe, a map probe, or a filler probe (4; only belief and map probes shown). After the participant's response, the video ends with the woman's return, and the participants are asked to indicate the location of the target object (5).

probes were unpredictable (Apperly et al., 2006; Cohen & German, 2009). We predicted that specialized processing mechanisms for mental states would result in a reaction time signature wherein belief probes would receive faster processing than map probes.

2. Experiment 1

2.1. Methods

Twenty-seven undergraduates (15 females and 12 males, $M = 19.1$ years, $SD = 1.41$ years) from the University of California, Santa Barbara participated for course credit. Participants watched videos of simple search action scenarios and were instructed to track the location of an object (Fig. 1). In each video, an actress put the object in a drawer and, before leaving, drew a map indicating its location to help a friend locate it. While gone, an officemate returned to put an unrelated object away in the room and either moved the target object ("false" conditions) or left it in place ("true" conditions). At this point a test probe presented with text (e.g., "She thinks that the purse is in the right drawer"; "The map shows that the purse is in the right drawer") or one of several different filler probes (e.g. "It is true that the purse swapped locations"; "It is true that she drew a map") interrupted the video.

The "y" and "n" keys were used for "yes" and "no" responses, respectively. The participant used his or her dominant hand which they were instructed to keep over the "h" key (which is in between the "y" key above it and the "n" key below it).

After participants gave a "yes" or "no" button-press response, the video finished and participants provided a second button-press response to indicate the object's true location, ensuring they followed instructions and tracked the object.

Videos ranged from 55 to 60 s and were presented with E-Prime 2.0 (Psychology Software Tools, Inc.) software. Responses to probes were equally likely to be "yes" or "no" and the object was equally likely to start off on the left or right side and to end up on the left or right side at the time of the probe.

Probes fell into four conditions: false belief, false map, true belief, and true map. Accuracy and reaction times (RTs) to probes were measured. Trials were randomly presented in a repeated-measures design, with 44 trials divided over four blocks (32 test trials and 12 filler trials).

2.2. Results and discussion

Reaction times[3] were subjected to a two-factor repeated-measures analysis of variance, revealing a main effect of representation type, $F_{(1,26)} = 19.5$, $p = .0002$, $\eta_p^2 = .429$ such that RT to beliefs were significantly faster than RT to maps, and a main effect of truth value, $F_{(1,26)} = 9.8$, $p = .004$, $\eta_p^2 = .274$ along with a significant interaction between representation type and truth value, $F_{(1,26)} = 4.46$, $p = .045$, $\eta_p^2 = .146$.

Inspection of Fig. 2 suggests this interaction results from the difference between beliefs and maps being considerably larger for false representations than for true. Pairwise post hoc t-tests confirmed that the comparison between false beliefs ($M = 2197$ ms, $SD = 389$ ms) and false maps ($M = 2429$ ms, $SD = 583$ ms) was significant, $t_{(26)} = 4.47$, $p = .0001$, $d = 0.86$, with an effect size twice as

[3] For all analyses, outlier trials (defined as ±3SD from the mean RT for each subject) and error trials (which were rare, with no indication of any speed-accuracy tradeoff) were excluded from the main analyses. An error analysis revealed that the number of incorrect responses did not significantly differ across trials. In the false belief, false map, true belief, true map conditions there were errors on 4%, 6%, 3%, and 3% of the trials, respectively. All p's > .05 for the ANOVA and pair wise comparisons.

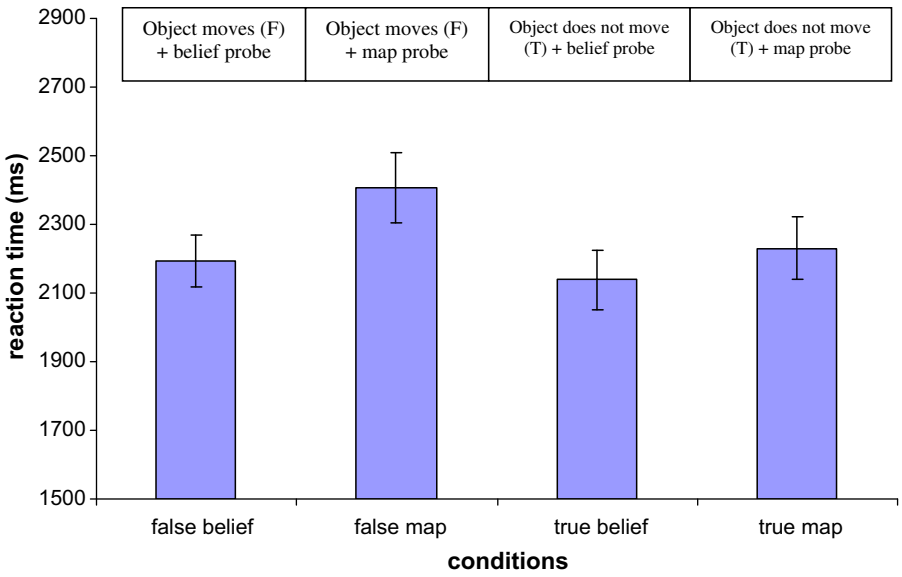

Fig. 2. Shows reaction times and standard errors for each of the four conditions (from left to right: false belief, false map, true belief and true map).

large as that obtaining from the difference between true beliefs (M = 2148 ms, SD = 462 ms) and true maps (M = 2224 ms, SD = 471 ms), $t_{(26)} = 2.18$, $p = .038$, $d = 0.42$. There was also a difference between false and true maps, $t_{(26)} = 3.30$, $p = .003$, $d = 0.63$, but no difference between false belief and true belief, $t_{(26)} = 1.42$, $p = .17$, $d = 0.27$.

Belief probes were three characters shorter than map probes so an additional analysis was conducted to rule out the possibility that the RT advantage for belief over map probes might be due to a difference in probe length. In this analysis (see, e.g. Trueswell, Tanenhaus, & Garnsey, 1994), actual RTs on all trials, including filler trials, and character length were used to generate a regression equation which in turn was used to compute estimated RTs for each subject for each test trial sentence.

Residual RTs were calculated as the difference between the predicted and actual RTs for each test trial, and the summary statistics for these residuals, along with RTs for each condition appear in Appendix A. An analysis on the residual RTs, which therefore controls for differences in probe length, revealed the same pattern of performance with the exception that the RT difference between true belief and true map conditions fell short of significance, $t_{(26)} = 1.04$, $p = .31$.

Despite no overt instructions to track the contents of either beliefs or maps, participants responded faster to unpredictable belief probes than they did to unpredictable map probes, a result consistent with the existence of domain specific systems specialized for processing mental state representations. When the estimated reading times for probe sentences was controlled for, the difference was confined to representations that were false in content. This might stem from the possibility that responses in the true content conditions are more closely based on the information that subjects' were explicitly instructed to track (i.e. the real location of the object).

The advantage for responding to probes about false beliefs over probes about false maps is consistent with the idea that there might be domain specialized mechanisms for mental state representations. However, a possible objection to consider stems from the wide variety of possible public representations, and the possibility that maps might be one type of public representation, that just happen to be a 'more complex' kind of public representation to represent and reason about than other kinds of public representations. For example, there are many different specific forms a map might take in order to specify the location of an object (e.g. different details might be made explicit, different notations might be used, etc.).[4] While this is true for beliefs as well, it is not clear that people need to consider the format of mental states in order to reason about them.

Prior to concluding that there is a general advantage for 'mental state' over 'public' representations, it would be desirable to show that an advantage exists even when beliefs are compared to another kind of public representation; ideally, a kind of public representation that is more constrained in the way that it conveys its content, and therefore may place lower demands on participants.

In Experiment 2, therefore, we compare participants' responses to belief probes with responses to probes about arrows, a form of public representation with a more constrained format than a map has.

3. Experiment 2

3.1. Methods

Twenty-five undergraduates (16 females and 9 males, M = 20.0 years, SD = 2.26 years) from the University of Cal-

[4] We are grateful to a reviewer for pointing out this possibility.

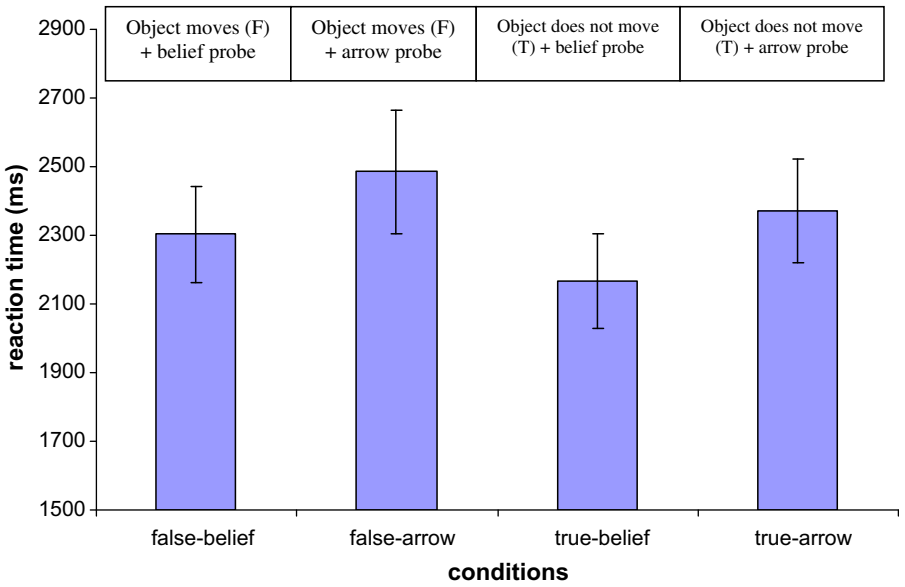

Fig. 3. Reaction times and standard errors, for each of the four conditions of Experiment 2 (from left to right, false belief, false arrow, true belief and true arrow).

ifornia, Santa Barbara participated for course credit. The design was identical to experiment 1. The procedure was also identical except participants received arrow test probes instead of maps (e.g., "The arrow shows the purse is in the right drawer").

3.2. Results and discussion

RTs[5] were subjected to a two-factor repeated-measures analysis of variance, revealing a main effect of representation type, $F_{(1,24)} = 9.74$, $p = .005$, $\eta_p^2 = .289$ such that RT to beliefs were significantly faster than RT to arrows, and a main effect of truth value, $F_{(1,24)} = 8.71$, $p = .007$, $\eta_p^2 = .266$ such that RTs to true representations were significantly faster than RTs to false representations. There was no interaction between representation type and truth value, $F_{(1,24)} = 0.08$, $p = .78$, $\eta_p^2 = .003$[6] (see Fig. 3).

Pairwise post hoc *t*-tests confirmed that the advantage for false beliefs ($M = 2302$ ms, $SD = 703$ ms) over false arrows ($M = 2485$ ms, $SD = 894$ ms) was significant, $t_{(24)} = 2.33$, $p = .03$, $d = 0.47$, and that RTs were also significantly faster for true beliefs ($M = 2167$ ms, $SD = 691$ ms) than for true arrows ($M = 2371$ ms, $SD = 752$ ms), $t_{(24)} = 3.01$, $p = .006$, $d = 0.60$. While true beliefs were calculated faster than false beliefs, $t_{(24)} = 3.20$, $p = .004$, $d = 0.64$, the difference between true and false arrows fell short of significance, $t_{(24)} = 1.63$, $p = .12$, $d = 0.33$.

Two slight differences between the results of Experiments 1 and 2 emerged. First, while there was a difference between true and false maps in Experiment 1, there was no equivalent difference between true and false arrows in Experiment 2. Second, the difference between true and false beliefs evident in Experiment 2 was not seen in Experiment 1.

It is not entirely clear why these differences emerged, though in the former case we speculate that one possibility may reflect a difference between the way in which maps and arrows function as public representations. Maps are intentionally created public representations that most often carry their content via some isomorphism with the real world; they are in a sense 'copies' of spatial information in the world. Arrows, on the other hand, carry no *intrinsic content*; they function as public representations only via the spatial orientation they have with respect to the world (i.e. where they are pointing). To illustrate this more concretely, consider that if the physical orientation of a map changes, the content it conveys remains the same, while if the orientation of an arrow changes, the information it conveys changes with it.

Because of this difference, and because of their initial tight relationship to reality, calculating the content of a map that is 'true' might make relatively low demands compared to calculating the content of a map that is false. With arrows on the other hand, since they are not true or false in virtue of intrinsic content, there may be no particular advantage in the case where they are true. The difference between the Experiments 1 and 2 here in the relative cost of processing true versus false content for each public representation type is entirely consistent with this interpretation, though further evidence on the question is required before a definitive answer can be offered.

[5] A difference in error rates between belief and arrow probes reached near significance, $p = .06$, however, this was in the opposite direction of that predicted by a speed-accuracy trade-off. All other effects and comparisons were not significant, $p > .05$. Error rates came in at 3%, 8%, 5%, and 9% for false belief, false arrow, true belief, and true arrow conditions, respectively.

[6] The number of words and characters for the belief and arrow probes were matched in Experiment 2. Nonetheless, a residual reaction time analysis was conducted as for Experiment 1 which revealed the same pattern of results as for the analysis reported here using raw RTs. Summary statistics for the residual analysis are presented in the Appendix.

The finding that true beliefs are calculated more quickly than false beliefs in Experiment 2 is more consistent with the existing proposed models of belief-desire reasoning (e.g. Leslie, German, & Polizzi, 2005) and other evidence from reasoning tasks with both younger and older adults (e.g. German & Hehman, 2006) showing response time advantages in calculating true beliefs over false beliefs. The lack of such a difference between false and true beliefs in Experiment 1 is thus the unexpected result here.

4. General discussion

Across two experiments, participants responded to unpredictable probes about the content of an agent's belief about the location of an object faster than unpredictable probes about the contents of a map (Experiment 1) or arrow (Experiment 2) about the location of an object.

What are the possible sources, in terms of cognitive architecture or processing, for the advantage for calculation of representations in the mental state domain over artificial representations in these two varying formats? We identify three related possibilities here.

First, the advantage might stem from accumulated 'expertise' with making calculations in the social domain based on greater experience with beliefs than with artificial representations (for which there is presumed lesser experience). Second, the advantage might stem from architectural specialization that exists in core mechanisms that are responsible for the initial acquisition of theory of mind knowledge (e.g. Leslie et al., 2004). Third, the advantage might stem from an inherent difference in the computational requirements for representing mental and public representations. These proposals, while differing in certain details, all commit to some version of domain specificity in the adult processing system. We deal with each of these proposals in turn.

Appeals to 'expertise' or 'amount of experience' with beliefs versus other kinds of representations are often made to obviate the need to propose architectural domain specificity for mental state content in the cognitive system (e.g. Perner, Ruffman, & Leekam, 1994). However, under such proposals, experience with mental state representations must at least *result* in domain specialization for that content type (and not others), even if specialization is assumed not to exist as a result of the core architecture of the system. Yet, if a domain general cognitive system has an architecture that *does not have the capacity to tell apart the two types of content* (which would be the strongest version of a domain general architecture), it is unclear why 'practice' with one kind of content (e.g. beliefs) would not generalize to other kinds of representations, and *vice versa*.

A second version of this domain generality might therefore propose instead that while processing *mechanisms* are domain general, an advantage for beliefs stems from belief content somehow having priority access to domain general mechanisms. On this kind of view, however, there is still domain specificity somewhere in the system; it is just located earlier in the processing stream, where belief content must have been 'sorted' somehow from other kinds of representations and allowed priority access to the domain general processor for representations.

It is worth also considering closely the premise driving expertise proposals; that humans have *more experience* with mental as opposed to public representations. While this is an *intuitively* plausible premise, the precise nature of how to quantify experience and its capacity to generate learning and expertise needs specification for one to make sense of 'expertise' proposals. For example, one obvious difference between 'public' representations and mental state representations is precisely that the contents of public representations such as maps, signs, photos and arrows are *public*, and thus in cases where their content matches the true state of affairs, this match is observable. Likewise, when the content of a public representation deviates from the true state of affairs, the mismatch is observable. Cases of 'misrepresentation' might therefore readily lend themselves to learning mechanisms that operate over such observable discrepancies (see e.g. Leslie, 2000, for a proposal related to this for how children might come to learn about hidden mental representations by analogy to public representations).

By contrast, mental states are not directly observable entities in the world, and therefore it is not obvious how evaluating 'matches' and 'mismatches' with observable behavior can drive learning. Cases where predicted actions mismatch an observed action might be salient experiences, but even so there are still multiple sources of variance between the action and the predicted mental states that might have caused the 'error' in prediction. For example, if an agent searches in an unexpected location, this might be because her belief was actually false, but it might instead have been because her desire changed instead (see Wertz & German, 2007, for evidence of adults offering multiple desire representations in action explanation).

Making the contents of one's mental states public, as one does in mental state related conversation, might play a role in facilitating such learning (see e.g. proposals such as that of Taumoepeau and Ruffman (2008)), but the details of exactly how such input is translated into increased expertise at reasoning about mental state contents are still obscure. So while the notion of differential experience with different types of content is intuitively appealing, there is yet work to do in specifying both how experience is to be quantified and how it drives expertise acquisition.

Theories that propose early specialized core architecture that includes domain-specific components (e.g. Leslie et al., 2004) gain some support from developmental neuropsychological investigations showing that children with autism are selectively impaired on 'theory of mind' tasks but are at ceiling on tasks involving non-mental representations (Leslie & Thaiss, 1992). This selective impairment persists in spite of the fact that the learning environment for children with autism is not obviously different than the learning environment of children without autism.[7]

[7] While differences in patterns of 'social attention' suggest that children with autism might attend differently to aspects of the social world and therefore experience it differently (Dalton et al., 2005; Klin et al., 1999; Ristic et al., 2005), some specialized attention patterns (e.g. 'agency monitoring', New, Cosmides, & Tooby, 2007) appear to be intact in autism (New et al., 2010).

Further evidence consistent with an early, reliably developing system for mental state inferences is provided by infants' expectations about sequences of events where an agent has a true or false belief – a sensitivity emerging no later than 15 months, suggesting that core mental state reasoning systems are in place early in development (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007).

However, one challenge to theories proposing core architectural domain specificity for theory of mind is recent evidence from brain injured patients, showing an association between performance on false belief and false photographs in patients with lTPJ lesions; such patients are impaired on both kinds of representational content (Apperly et al., 2007).

While association in impairment in these patients sits more easily with the idea of shared processing across content types, there are other explanations of this finding that remain consistent with core architectural domain specificity. Note that fMRI evidence points to rTPJ as the area most selectively activated when beliefs are compared to false photographs, with lTPJ showing a wider response profile (Perner et al., 2006). It is therefore possible that a more selective impairment with belief content might manifest in patients with damage to TPJ on the right side (a patient group that to date has not been tested, so far as we are aware). Second, and also consistent with core architectural functional specialization, is that two distinct neural subpopulations might be physically interleaved in the same anatomical area. Damage to lTPJ might compromise both of these distinct neural subpopulations, impairing performance across tasks that are nonetheless functionally distinct (see e.g. Kanwisher, 2000, for a similar argument about domain specific face perception in the right fusiform gyrus).[8]

A final possible account for the advantage of false belief content over false maps and false arrows is intermediate between strong domain specificity and strong domain generality. On this third approach, domain specialized circuitry that has the primary function of handling mental state content might be co-opted into processing public representations (e.g., arrows, maps and signs), owing to sufficient overlap in the problem content and format of the representations involved (see e.g. Barrett, 2005).

The idea is that mechanisms that form part of the core architecture for mental state reasoning, nonetheless also process non-mental representational content, albeit with less efficiency, which would explain both the advantage for the 'proprietary' content of the target domain (e.g. beliefs) over the extended domain (other public representations) and also the association in performance seen in the impairment in patients with damage to lTPJ (Apperly et al., 2007).

One possible inherent difference between mental state representations and public representations might lie in the computational requirements that are entailed in each case. In virtue of their being public, artificial representations such as maps and arrows exist in an explicit and observable medium. That is, we know when we reason about the content of a public artificial representation *how* the representation is conveying its content. An arrow pointing to the left drawer, and a map depicting the object in the left drawer both express the same content, but differ in the details of the format and are thus different representations. While beliefs, if they in fact exist at all (Churchland, 1981), must exist in *some* format, reasoning about someone's belief does not require one to know or assume anything about that format (see e.g. Leslie, 2000, for further discussion of this idea).

It is very unlikely that we consider the *format* or *medium* carrying belief contents (e.g. whether the belief exists as a "sentence in the head", or a "picture in the head") when we reason about mental states in everyday action prediction; we would consider two beliefs expressing the same content in different ways as *the same belief*. However, it is possible that reasoning about public representations, because of their very publicity, mandates the representational format to be specified by whatever cognitive architecture handles the reasoning. The requirement to make format explicit may be a computational task that is not routinely required in mental state reasoning, but that is a mandatory additional step in the processing of public representations. A system specialized for dealing with mental states in terms of brute propositional content (e.g. beliefs) might behave less efficiently if it has to deal with a representation in which the format must be made explicit, and this could explain the performance difference in processing beliefs and other public representations, even if the same architecture is employed for both tasks.

This third possibility, if true, would predict that the complexity of the format of the public representation might contribute to the efficiency with which it is processed. Simpler formats ought to be processed more efficiently than more complex formats. The current evidence across Experiments 1 (maps) and 2 (arrows) did not suggest any sizeable difference in processing speeds for the two different representational types, despite an intuition that arrow representations have a simpler format and therefore would be processed more quickly.[9]

A full exploration of this proposal requires further evidence, and pending such evidence, we tentatively attribute the advantage seen for belief content over public content observed across the experiments here as evidence for models proposing that mental state reasoning comprises (at

[8] This possibility would not be without precedent in the domain of theory of mind. Exogenous orienting of attention based on centrally presented directional cues also has its basis in rTPJ, leading researchers to question the specificity of the region for 'theory of mind' (Mitchell, 2008). However, high resolution fMRI comparisons of peak activation on the two tasks may also be consistent with functionally distinct neural populations being involved (Scholz et al., 2009).

[9] A two-way mixed design ANOVA comparing maps from experiment 1 and arrows from experiment 2 confirmed there was no main effect of representation type, $F_{(1,50)} = .299$, $p = .587$, $\eta_p^2 = .006$, a main effect of truth value, $F_{(1,50)} = 12.4$, $p = .001$, $\eta_p^2 = .199$, and no interaction between representation type and truth value, $F_{(1,50)} = 1.04$, $p = .312$, $\eta_p^2 = .020$. As expected, neither the difference between false maps ($M = 2429$ ms, $SD = 583$ ms) and false arrows ($M = 2485$ ms, $SD = 894$ ms) was significant, $t_{(50)} = 0.266$, $p = .791$, nor was the difference between true maps ($M = 2224$ ms, $SD = 471$ ms) and true arrows ($M = 2371$ ms, $SD = 752$ ms), $t_{(50)} = 0.855$, $p = .397$.

Table A1
Actual and 'residual' mean reaction times in experiment 1.

	False belief	False map	True belief	True map
RT means (SE)	2196.85 (74.83)	2429.24 (112.27)	2148.3 (88.91)	2223.54 (90.61)
Residual RT means (SE)	−40.11 (34.16)	161.05 (44.41)	−88.47 (27.86)	−44.04 (28.91)

Table A2
Actual and 'residual' mean reaction times in experiment 2.

	False belief	False arrow	True belief	True arrow
RT means (SE)	2302.44 (140.6)	2484.59 (178.83)	2166.86 (138.19)	2371.17 (150.48)
Residual RT means (SE)	−25.35 (36.6)	157.23 (51.7)	−159.98 (36.64)	44.64 (42.02)

least in part) of core architectural domain specific mechanisms underwritten by dedicated brain circuitry (such as rTPJ). In response to certain input conditions these mechanisms facilitate fast, automatic (or at least spontaneous) encoding of, and inferences about, agents' mental states (e.g. Cohen & German, 2009).

Acknowledgments

We would like to thank Daniel Bernstein, and three anonymous reviewers for comments on a previous draft of this manuscript.

Appendix A

Tables A1 and A2.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews, Neuroscience, 7*, 268–277.
Apperly, I. A., Riggs, K. J., Simpson, A., Samson, D., & Chiavarino, C. (2006). Is belief reasoning automatic? *Psychological Science, 17*, 841–844.
Apperly, I., Samson, D., Chiaravino, C., Bickerton, W. L., & Humphreys, G. W. (2007). Testing the domain specificity of a theory of mind deficit in brain injured patients: Evidence for consistent performance on non-verbal, "reality unknown" false belief and false photograph tasks. *Cognition, 103*, 300–321.
Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporoparietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16*, 1773–1784.
Bach, L. J., Happé, F., Fleminger, S., & Powell, J. (2000). Theory of mind: Independence of executive function and the role of the frontal cortex in acquired brain injury. *Cognitive Neuropsychiatry, 5*, 175–192.
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. (2005). Enzymatic computation and cognitive modularity. *Mind and Language, 20*, 259–287.
Bird, C. M., Castelli, F., Malik, O., Frith, U., & Husain, M. (2004). The impact of extensive medial frontal lobe damage on 'theory of mind' and cognition. *Brain, 127*, 914–928.
Bloom, P., & German, T. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, 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*, 259–283.
Charman, T., & Baron-Cohen, S. (1992). Understanding beliefs and drawings: A further test of the metarepresentation theory of autism. *Journal of Child Psychology and Psychiatry, 33*, 1105–1112.
Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67–90.
Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition*, 356–363.
Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*, 519–526.
Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., et al. (1995). Other minds in the brain: A functional imaging study of 'theory of mind' in story comprehension. *Cognition, 57*, 109–128.
Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 319–329.
Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*, 531–534.
Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study on 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*, 11–21.
Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*, 814–821.
German, T., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*, 129–152.
German, T., Niehaus, J. L., Roarty, M. P., Giesbrecht, B., & Miller, M. B. (2004). Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *Journal of Cognitive Neuroscience, 16*, 1805–1817.
Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759–763.
Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science, 11*, 159–165.
Klin, A., Sparrow, S. S., de Bildt, A., Cicchetii, D. V., Cohen, D. J., & Volkmar, F. R. (1999). A normed study of face recognition in autism and related disorders. *Journal of Autism and Developmental Disorders, 6*, 499–508.
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. (2000). How to acquire a representational theory of mind. In D. Sperber (Ed.), *Metarepresentations: An multidisciplinary perspective* (pp. 197–223). Oxford: Oxford University Press.
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., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*, 45–85.
Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.
Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory of mind. *Cerebral Cortex, 18*, 262.
New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences, 104*, 16593–16603.
New, J. J., Schultz, R. T., Wolf, J., Niehaus, J. L., Klin, A., German, T. C., et al. (2010). The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. *Neuropsychologia, 48*, 51–59.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
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*, 245–258.
Perner, J., & Leekham, S. (2008). The curios incident of the photo that was accused of being false: Issues of domain specificity in development, autism and brain imaging. *Quarterly Journal of Experimental Psychology, 61*, 76–89.
Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious; you catch it from your sibs. *Child Development, 65*, 1224–1234.
Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review, 9*, 507–513.

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research, 24*, 715–718.

Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain, 124*, 600–616.

Russell, J. (1999). Cognitive development as an executive process - in part: A homeopathic dose of Piaget. *Developmental Science, 2*, 247–270.

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

Saxe, R., & Powell, J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science, 17*, 692–699.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*, 1391–1399.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*.

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

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*, 640–656.

Stuss, D. T., Gallop, G. G., Jr., & Alexander, M. P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain, 124*, 279–286.

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

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24 and 33 months. *Child Development, 79*, 284–302.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.

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

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

Yazdi, A. A., German, T., Defeyter, M. A., & Siegal, M. (2006). Competence and performance in belief-desire reasoning across two cultures: The truth, the whole truth and nothing but the truth about false belief? *Cognition, 100*, 343–368.

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