COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13364

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13364

Evidence for a Weak but Reliable Processing Advantage for False Beliefs Over Similar Nonmental States in Adults

Steven Samuel,[a,b] Geoff G. Cole,[c] Madeline J. Eacott,[c] Rebecca Edwardson,[b] Hattie Course[b]

[a]*Department of Psychology, School of Health & Psychological Sciences, City University of London*
[b]*School of Psychology, University of Plymouth*
[c]*Department of Psychology, University of Essex*

Received 3 October 2022; received in revised form 27 May 2023; accepted 19 September 2023

Abstract

The ability to understand the mental states of others has sometimes been attributed to a domain-specific mechanism which privileges the processing of these states over similar but nonmental representations. If correct, then others' beliefs should be processed more efficiently than similar information contained within nonmental states. We tested this by examining whether adults would be faster to process others' false beliefs than equivalent "false" photos. Additionally, we tested whether they would be faster to process others' *true* beliefs about something than their own (matched) personal knowledge about the same event. Across four experiments, we found a small but reliable effect in favor of the first prediction, but no evidence for the second. Results are consistent with accounts positing specialized processes for (false) mental states. The size of the effect does, however, suggest that alternative explanations such as practice effects cannot be ruled out.

Keywords: Theory of Mind; False belief; False photo; True belief; Domain specialization

Humans often need to generate representations, things in our brains which "stand for" or code something in the world. For example, we will often need to represent another's mental state when we want to understand or predict their behavior, an ability often referred to as

Correspondence should be sent to Steven Samuel, Department of Psychology, School of Health & Psychological Sciences, City University of London, Northampton Square, London, EC1V 0HB, UK. E-mail: steven.samuel@city.ac.uk

Theory of Mind (ToM: Premack & Woodruff, 1978). In the classic false belief task (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983), an agent (e.g., "Sally") hides an object in one location only for another person to move the object to a different location without her knowledge. To reason about where Sally now believes the location of that object to be, a mental representation of her (false) belief must be generated.

A key debate about ToM is whether it is served by a specialized mechanism (Baron-Cohen, 1995; Cohen, Sasaki, & German, 2015; Leslie, Friedman, & German, 2004; Wellman, 2018). Historically, this has been explored by asking participants not only where Sally believes the object to be but also where a photo, taken before the object was moved, would show it (Zaitchik, 1990). Since the correct answer is the same for both questions, earlier or better performance on the belief question could support such a hypothesis. Using this paradigm, Leslie and Thaiss (1992) found that autistic children have difficulty solving false belief tasks but perform similarly to typically developing children when the problem concerns what a photo or map will show, leading them to conclude that there is indeed a specialized mechanism for belief reasoning (and one that is impaired in autism). However, the question of a ToM deficit in autism is debated (Brewer et al., 2016; Iao & Leekam, 2014; Milton, 2012; Williams, 2021), and although evidence from selective impairments is informative any specialization should also manifest in typical processing.

Some studies from the last decade or so have indeed found that adults make judgments about beliefs more efficiently than judgments about photo, text, note, or map equivalents (Cohen & German, 2010; Cohen et al., 2015; Samuel, Durdevic, Legg, Lurz, & Clayton, 2019). However, the problem with this evidence is that it comes from tasks that permit alternative explanations. For example, Samuel et al. (2019) asked participants to perform a video false belief task in which they watched short clips of a woman putting a gift in a drawer and then taking a photo of it before leaving. A man then entered, removed the gift, and either replaced it where he found it (true belief/true photo trials) or moved it to the other drawer in the desk (false belief/false photo trials). An icon then appeared which prompted participants to press a button to indicate which drawer the woman believes the object to be in or where the photo will show it to be. In support of more efficient processing of mental states, adults made decisions about false beliefs on average more quickly than they did about matched photos or text messages. However, every trial was performed in dual-task conditions in which participants had to hold in mind a sequence of numbers or a visual pattern. The extra difficulty produced by the secondary task may have disproportionately impacted nonmental state trials because they are less practiced (i.e., people have less experience thinking about what photos show and texts say than what people think). This leaves open the possibility that the reason false belief trials were performed more quickly is because they were less susceptible to dual-task interference, not that there is a specialized process for mental state reasoning.

In an earlier and similar study by Cohen et al. (2015) there was no secondary task, but instead of being prompted to respond by icons participants verified sentences of the type *she thinks that the purse is in the left drawer* (mental state) and *the note shows that the purse is in the left drawer* (nonmental state). Again, in support of more efficient mental state processing, responses were faster on false belief trials than on false note trials. The absence of a reliable difference on *true* belief/note trials rules out the possibility that something about the sentence structure accounted for this effect. It could, however, be that the *falseness* in *she thinks* sentences is processed more quickly than in the *note shows* sentences because, again, adults are more used to language concerning things that people (falsely) think. Thus whereas a potential issue with the study by Samuel et al. (2019) is that dual-task interference could account for the more efficient processing of mental states, the results of Cohen and colleagues may reflect differences in the efficiency of how meaning is processed.

Although contrasting mental and nonmental representations has been a popular means of testing domain specificity, it is perhaps not the only way; mental state processing can also be contrasted with personal (declarative) knowledge, which does not require representation because it is the participant's own information. For example, if participants were faster to respond to a question about another agent's true belief about the location of an object than even a direct question to the participant about where the object *is*, then the case for domain specificity would be strengthened. There is an additional benefit to investigating true beliefs; since the majority of beliefs *are* likely to be true (Dennett, 1989; Leslie, 1994), the prevailing focus on false beliefs effectively excludes the greater proportion of our ability (Saxe, Carey, & Kanwisher, 2004). However, researchers traditionally treat true belief trials as a baseline or filler trial rather than a dependent measure (e.g., Apperly, Samson, Chiavarino, Bickerton, & Humphreys, 2007; Hale & Tager-Flusberg, 2003; Samuel et al., 2019) because of the "true belief problem." As highlighted by Dennett (1978), it is impossible to know whether participants actually consider others' beliefs rather than use reality itself to obtain an answer, since both routes arrive at the same conclusion.

It is important to highlight that no theory of domain specificity in ToM currently posits that others' true beliefs should ever be processed more efficiently than our own knowledge, and a domain-specific account would not be prejudiced by the absence of such a finding. However, evidence in favor of more efficient processing of beliefs would imply particularly convincing support for domain specificity, even though the converse would not be to its detriment. There are also empirical reasons to believe that the true belief versus knowledge contrast might be informative. If the delay between the encoding of a belief and the assessment of that belief is short (3 s as opposed to 23 s), participants have been shown to be faster to verify statements about others' true beliefs such as *She thinks that the object is on the right* than statements of fact like *It is true that the object is on the right* (Cohen & German, 2009). Again, this would appear to support the case for domain specificity, but an important caveat to this finding is that the belief statement begins with "She," which should prime participants to think about the woman's role earlier than the relatively vague "it" in the latter sentence. Moreover, the researchers themselves did not predict this advantage for beliefs, and did not interpret these specific data as evidence for domain specialization.

1. The present experiments

The current experiments have two main aims. First, we examined whether the speed advantage for beliefs over nonmental states holds when prompts are icons rather than sentences and when there is no secondary task interference. This should rule out the two potential confounds

Table 1
Mean response times (ms) and standard deviations

	Knowledge		Belief		Photo	
Exp	Affirmative	Negative	True	False	True	"False"
2019	1334 (495)	*Not analyzed*	1351 (382)	1380 (279)	*Not analyzed*	1551 (297)
1	957 (504)	1268 (575)	1052 (548)	1095 (557)	1005 (518)	1161 (518)
2	734 (227)	868 (285)	779 (253)	924 (276)	785 (232)	926 (250)
3	849 (235)	1060 (392)	868 (304)	1003 (288)	905 (301)	1081 (305)
4	1147 (364)	1151 (282)	1039 (341)	1116 (345)	1026 (356)	1219 (365)
Combined	924 (380)	1087 (422)	936 (390)	1035 (387)	931 (376)	1098 (381)

Note. Note that response times for affirmative knowledge are based only on those trials in which the gift did not change location (so as to match true belief trials).

described in relation to the studies by Cohen et al. (2015) and Samuel et al. (2019). Second, we assessed whether judgments about others' true beliefs would be faster than judgments based on facts (henceforth termed *affirmative knowledge* trials). The first of these predictions follows from the domain-specific hypothesis; the latter is more exploratory given the caveats discussed above.

Prior to collecting data, we performed new analyses of the data from the study by Samuel et al. (2019) in order to establish a benchmark for preparing the present experiments (see Supplementary Materials for details). In this earlier study (described briefly above), participants performed a concurrent interference task, and, therefore, results were considered merely indicative of any "true" effect sizes of the measured differences. Descriptive statistics and results are displayed in Tables 1–3. False belief trials were performed significantly faster than false photo trials with an effect size of the contrast close to large, and a Bayesian analysis suggesting the data were 20 times more likely under the alternative hypothesis. However, there was no evidence of a difference in response times (RTs) between true belief and affirmative knowledge trials (a negligible effect size and the Bayesian analysis clearly favored the null hypothesis). These reanalyses suggest on the one hand that mental state processing may be substantially more efficient than nonmental state processing, but that reasoning about true beliefs may not differ from assessing direct knowledge. The following experiments test these possibilities without the confounds discussed above.

2. All experiments: General method

The preregistration documents for each experiment can be found here: Exp 1: https://osf.io/prj67; Exp 2: https://osf.io/6fbpv; Exp 3: https://osf.io/cdk9a; Exp 4: https://osf.io/qszj2.

Table 2
Mean accuracy (%) and standard deviations

	Knowledge		Belief		Photo	
Exp	Affirmative	Negative	True	False	True	"False"
2019	97	*Not analyzed*	98	97	*Not analyzed*	96
	(7)		(5)	(6)		(14)
1	99	96	96	96	97	95
	(2)	(6)	(11)	(8)	(7)	(7)
2	96	84	96	92	93	89
	(10)	(21)	(9)	(13)	(12)	(22)
3	99	92	92	94	99	92
	(5)	(18)	(16)	(11)	(4)	(15)
4	97	94	95	89	96	87
	(7)	(12)	(10)	(19)	(10)	(24)

Note. Note that accuracy for affirmative knowledge is based only on those trials in which the gift did not change location (so as to match true belief trials).

2.1. Participants

A target of 34 participants was set for each experiment, a number calculated using G*Power 3.1 for an 80% chance of detecting medium effect sizes ($d = .5$) on two-tailed matched-pairs t-tests. Participants were recruited from UK universities and required to be aged between 18 and 35 (Exp 1: $N = 34$, $M_{Age} = 19$ years, range 18–27, 5 men, 29 women; Exp 2: $N = 34$, $M_{Age} = 20$ years, range 18–34, 8 men, 26 women; Exp 3: $N = 33$, $M_{Age} = 19$ years, range 18–27, 16 men, 17 women; Exp 4: $N = 35$, $M_{Age} = 19$ years, range 18–30, 1 man, 34 women). In Experiment 3 (only), participants were required to have English as a first language, as this experiment also included the Autism Quotient questionnaire in an exploratory investigation. All received course credit.

3. Method

3.1. All experiments

Participants were informed that they were going to see some sped-up videos in which (precise instructions follow) "a woman hides a gift for a friend to discover the next time they return to their desk. She also takes a photo of the gift in its drawer to send to another friend, who also contributed to the gift.[1] She then closes the drawer and leaves the room. However, later that day, a male colleague who is NOT the friend in question comes into the room to look for an important piece of paper. While doing so, he has to take the gift out of the drawer briefly. Sometimes he puts it back in the same place he found it, but other times he is a bit absent-minded and puts it in a different drawer."

After instructions, participants performed 12 practice trials with feedback (green square for correct responses, red square for incorrect responses and timeouts), with two trials of

Table 3
Results of comparisons (mean RTs)

Exp	True belief ≠ Affirmative reality					False photo ≠ False belief				
	t (df)	95% CI_{MDiff}	p	d	BF_{10}	t (df)	95% CI_{Diff}	p	d	BF_{10}
2019	0.452 (46)	[−61, 96]	.65	0.066	0.17	3.521 (23)	[70, 271]	.002**	0.719	20.67
1	1.247 (33)	[−60, 251]	.22	0.214	0.37	1.887 (33)	[−5, 139]	.07	0.324	0.90
2	1.526 (33)	[−15, 106]	.14	0.262	0.53	0.044 (33)	[−95, 91]	.97	0.007	0.18
3	0.502 (32)	[−58, 96]	.62	0.087	0.21	2.191 (32)	[6, 151]	.036*	0.381	1.51
4	2.615 (34)	[23, 193]	.013*	0.442	3.38	2.417 (34)	[16, 190]	.021*	0.409	2.27
1–4	0.466 (135)	[−38, 62]	.64	0.024	0.11	3.114 (135)	[23, 103]	.002**	0.267	9.41

Note: *Significant at the $p < .05$ level, **Significant at the $p < .01$ level.

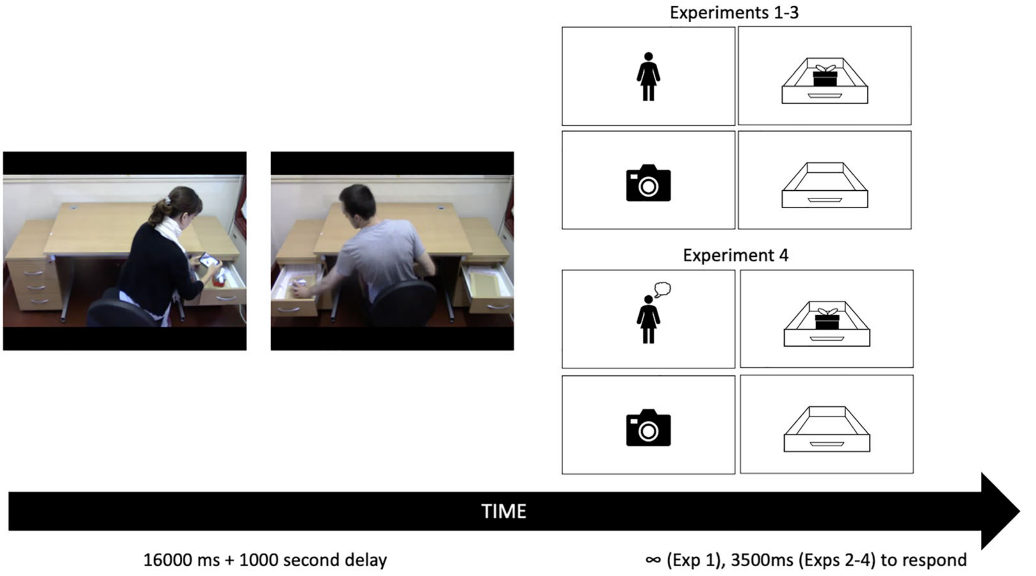

Fig. 1. Participants watch videos in which a woman places an object (a gift) in a drawer and takes a photo of it. Later, in the woman's absence, a man searches for something in the drawers and takes the gift out in the process. He either replaces the gift where he found it (true belief/true photo trials) or, as here, inadvertently puts it in the other drawer (false belief/false photo trials). When the video ends, participants are shown a cue which indicates whether they should respond (make a left or right button press corresponding to the left and right drawer) according to the woman's belief about where the gift is, where the photo will show the gift to be, where the gift actually is now, or which drawer does not contain the gift.

each type (true belief, true photo, false belief, false photo, affirmative knowledge, negative knowledge). Trial order was always randomized. Each trial began with a blue square which remained on screen until the participant pressed the *b* key on the keyboard to start the next trial. This action was carried out at the participant's own pace. Next, the video was shown (16 s), followed by a 1-s delay prior to the presentation of the cue. The trial procedure is illustrated in Fig. 1, and an example video is available here (https://www.youtube.com/watch?v=np6-l2q-EP4). In every video, a woman places a gift in one of two drawers in a desk (left or right), then takes a photo of it with her phone. After she leaves the scene, a man enters, takes the gift out of its drawer, then searches the drawers as if looking for something. He then places the gift either in the original drawer or the only other drawer before he leaves. Participants are then cued by an icon as to how to respond. Specifically, for the woman icon, the instruction was always to "press the left or right key to indicate which drawer the woman believes the gift to be in" (belief trials), for the camera icon always "which drawer the photo will show the gift to be in" (photo trials), for the gift icon always "which drawer the gift is in now" (affirmative reality trials), and for the empty drawer icon always "which drawer does NOT contain the gift now" (negative knowledge trials). Participants then made a left (*v* key) or right (*n* key) button response corresponding to the left or right drawer, respectively. Instructions were to respond as quickly and as accurately as possible. Responses that were more than 2.5 times above the standard deviation of the mean for that trial type and participant were removed. Note that only those affirmative knowledge trials where the object *did not move* were used in analyses, not those where the object did move. This was necessary to match them with true belief trials (where the object also never moved). Starting gift position (left/right) and correct answer (left/right) was counterbalanced across the task as a whole and within each specific trial type. Like the practice trials, the order of experimental trials was randomized.

3.2. Experiment 1

The experiment was run at the University of Essex, UK, using ePrime 2.0 software. No time limit was applied to responses. After the 12 practice trials, participants performed 96 experimental trials in total, 12 each of true belief, false belief, true photo, and false photo. There were also 24 affirmative knowledge trials and 24 negative knowledge trials. This meant that each icon was shown an equivalent number of times. True belief trials were only compared to affirmative reality trials where the object did not move (half of the affirmative reality trials). For Experiment 1, 25 participants had at least one trial removed ($<1\%$ data) for exceeding the threshold of 2.5 standard deviations above the mean for that participant and trial type.

3.3. Experiments 2 and 3

The experiments were run at the University of Plymouth, UK, using Open Sesame software (Mathot, Schreij, & Theeuwes, 2012). They were designed to reduce fatigue (total experiment time reduced from approximately 45 to 30 min) and ensure response times were as fast as possible. After the 12 practice trials, participants performed 64 (not 96) experimental trials, and responses were now required within 3500 ms of cue presentation or the trial would time out. In both experiments, there were eight true belief, eight false belief, eight true photo, and eight false photo trials. A further 16 trials were affirmative knowledge trials, and 16 were negative knowledge trials. As before, true belief trials were only compared to affirmative reality trials where in the latter the object did not move (half of the affirmative reality trials). All RTs used in the analyses were within 2.5 times standard deviations above the mean for that trial type and participant and thus no removals were necessary. One participant in Experiment 2 and five in Experiment 3 did not meet the minimum 50% accuracy threshold and were replaced (final overall accuracy for each sample: Exp 1: $M = 96\%$, range 71–100%; Exp 2: $M = 91\%$, range 63–100%; Exp 3: $M = 94\%$, range 66–100%). In Experiment 3, we also gave people the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). This was to conduct an exploratory analysis of whether greater autistic-like traits were related with the ability to process false beliefs relative to false photos, and true beliefs relative to reality. Unfortunately, practical circumstances (expiration of researcher contracts) meant that only 33 of the 34 planned participants were recruited for Experiment 3, meaning this did not meet all the requirements set out in its preregistration.

3.4. Experiment 4

Experiment 4 was run at City, University of London (UK) and was identical to Experiment 2 apart from two changes. First, the icon to cue a response about the woman's belief was adapted to include a thought bubble. This was done because while the camera icon has a restricted range of potential interpretations, the woman icon is used and seen in multiple contexts. While the instructions participants were given were explicit that the icon meant they should make a response based on the woman's belief, adding the thought bubble should equate the two icons better on this variable and minimize the possibility that differences (or similarities) between trial types were due to this factor. Second, it was noticed that in Experiments 2 and 3, the backstory given to participants no longer mentioned that the friend who the female protagonist sends the photo to also contributed to the gift. While there was no obvious theoretical reason why this may have led participants to behave differently in the task, this was included here to make the instructions identical again to Experiment 1 and the original experiment (Samuel et al., 2019). For Experiment 4, two participants had one trial removed ($<1\%$ data) for exceeding the threshold of 2.5 standard deviations above the mean for that participant and trial type. As this experiment was run on multiple computers simultaneously, one extra participant's data were retained ($N = 35$ total).

As part of our preregistration for Experiment 4, we also added a combined analysis of the data from all four experiments presented here. Following the results from Experiments 1–3, it was hypothesized that there would be no difference between mean response times on false belief and false photo trials in this analysis.

4. Results

Descriptive statistics of response times and accuracy are displayed in Tables 1 and 2, and the results of the confirmatory tests are in Table 3.

4.1. False belief versus false photo

Confirmatory analyses found significant differences for Experiments 3 and 4, with false belief trials performed more quickly. However, no such difference was found in Experiments 1 or 2. Bayesian analyses did not show convincing support for the alternative hypothesis for either of the two statistically significant effects. Effect sizes ranged from negligible to small-to-medium at best. Overall, there was some support for more efficient processing of false beliefs than false photos.

4.2. Affirmative knowledge versus true belief

In the first experiment, no significant difference between affirmative knowledge and true belief trials was found and the latter were in fact the slower trial type numerically. Bayesian analyses favored this null result. When this pattern was revealed in Experiment 1, the preregistered hypothesis from Experiment 2 onward was switched, now with the expectation that true belief trials may be performed more slowly than affirmative reality trials, with the new hypothesis that *representations* (generally) may be processed more slowly than knowledge (this hypothesis could function alongside the possibility that false beliefs might be performed more quickly than false photos, as these concerned a comparison between two representations). However, Experiments 2 and 3 found no evidence for this revised hypothesis, and a (solitary) significant difference in Experiment 4 instead favored the *original* hypothesis that true beliefs would be processed more efficiently. This isolated result was also supported by a Bayesian analysis.

4.3. Relationships with Autism Quotient (Experiment 3 only)

There was no evidence for a relationship between AQ scores and the effect of reasoning about true beliefs over affirmative knowledge (mean true belief RT minus mean affirmative knowledge RT), rho $= -.06$, $p = .73$, or the effect of reasoning about false beliefs over false photos (mean false photo RT minus mean false belief RT), rho $= -.04$, $p = .82$.

4.4. Combined analysis (Experiments 1–4)

As part of our preregistration for Experiment 4, we also planned a combined analysis with the data from all four experiments. This increases the power of the statistical analysis but at the expense of failing to distinguish between any impact the subtle changes between experiment designs may have made.

Contrary to expectations when we preregistered the combined analysis, but consistent with our initial hypothesis, confirmatory analysis revealed a statistically significant effect, with false belief trials performed on average 63 ms faster than false photo trials, with a small effect size ($d = 0.267$). This result was also supported by a Bayesian analysis, which suggested the combined data were approximately nine times more likely under the alternative hypothesis (two-tailed). No significant difference was found between response times on true belief and affirmative reality trials. This was supported by a Bayesian analysis, and by the negligible effect size ($d = 0.024$).

4.5. Exploratory tests: Validity

Exploratory tests were also conducted to check that something about the *icons* representing beliefs and photos did not lead to variation in response times. This was important because it could be that people might be more practiced at processing icons of people than camera icons, and may, therefore, be faster on belief trials because the former icons are understood more quickly rather than because of any advantage in processing mental states. Contrary to this possibility, an analysis of the combined dataset found *true* belief and photo trials were processed with similar speed, $t(135) = 0.277$, $p = .78$, $d = 0.024$. It is worth noting that this was also the case for Experiment 4 alone, where the inclusion of a thought bubble meant the icon differed from Experiments 1–3, $t(34) = 0.344$, $p = .73$, $d = 0.058$, $BF_{10} = 0.10$.

Additionally, exploratory tests were conducted on the same combined dataset to test the assumption that response times on false beliefs and false photo trials would be slower than when these were true. This was also the case (true vs. false belief: $t(135) = 5.069$, $p < .001$,

$d = 0.435$, $BF_{10} = 9816$; true vs. false photo: $t(135) = 8.237$, $p < .001$, $d = 0.706$, $BF_{10} > 9999$). Together, these results suggest the task was appropriate for measuring differences in performance when reasoning about mental and nonmental states.

Overall, the combined data from all four experiments suggest a small but significant processing advantage for false beliefs over false photos, but no difference between the processing of true beliefs and reality itself.

5. General discussion

In a series of studies, we tested whether previous evidence of processing advantages for false beliefs over similar but nonmental representations was the result of the use of linguistic prompts and/or indirect effects of concurrent interference tasks. By removing these extraneous factors and focusing on behavioral performance rather than neurological data, the tasks presented here featured response times of around 1000–1100 ms, much faster than in previous research (over 2000 ms in Cohen et al., 2015; around 1400 ms in Samuel et al., 2019; over 2500 ms in Saxe & Kanwisher, 2003). This suggests that these tasks are well-placed to test the specialization hypothesis while keeping noise from extraneous factors to a minimum.

Overall, our results offered support for a weak but reliable advantage for false belief processing even in the absence of these extraneous factors. In the first three experiments, we found evidence in favor of faster false belief processing only once, but after the inclusion of a fourth experiment, a combined analysis revealed a significant difference overall, with a small effect size and support from Bayesian analyses. This small effect size could explain why statistically significant findings were inconsistent at the level of the individual experiments; they were designed to detect medium effect sizes. This small effect also makes it difficult to interpret the absence of any relationship with the AQ (Experiment 3), which, therefore, almost certainly required a much larger sample size to make a sensitive enough test. In contrast, there was no evidence for our (original or revised) secondary hypothesis; affirmative reality and true belief trials were processed with similar efficiency, an outcome also supported by Bayesian analyses.

The null results for the secondary hypothesis are of much less theoretical significance for our understanding of ToM than the positive result for the false belief versus false photo contrast, since domain-specialization theories make specific predictions only about the latter. Here, our results are consistent with such theories (Baron-Cohen, 1995; Cohen et al., 2015; Leslie et al., 2004). Even when secondary interference tasks and linguistic prompts are removed, participants were faster to respond when they were required to consider an agent's mental state rather than the contents of a photo, despite the result (and necessary left/right response) being identical. Importantly, this effect could not have been driven by familiarity with the woman icon over the camera icon because the effect did not emerge for the true belief versus true photo contrast.

However, when compared to other studies using a similar paradigm, it is noteworthy that the overall effect size of the false belief response time advantage was small. When contrasting false beliefs with false notes, false maps, and false arrows, previous experiments have generated effect sizes that are typically medium and even *large* (Cohen & German, 2010; Cohen et al., 2015). The previous study using the same paradigm as in the present studies also found a medium-to-large effect size difference between false beliefs and false photos, using exactly the same videos and icons as used here (Samuel et al., 2019). These studies also consistently found a statistically significant difference between the two trial types. Here, the results were more equivocal. One explanation for this discrepancy concerns statistical power; the studies by Cohen and colleagues involved around 70 participants each, while in the studies here, the samples were closer to 30. Once we combined the data, our results patterned as in earlier studies, although the effect size did not. Another explanation, perhaps complementary to the first, is that removing linguistic prompts and interference tasks may have reduced the magnitude of the processing advantage for false beliefs. This would suggest that even specialized processing of mental states is cognitively penetrable. Without an experiment to test this possibility directly (e.g., by comparing the magnitude of the effect when the prompts are either sentences or icons), this is necessarily speculative. Nevertheless, even a weakened effect would be consistent with accounts that posit theory of mind mechanisms which allow for some cognitive penetrability from other factors. Although it is not made explicit by its originators, this would seem to be permitted in the by-product account, whereby a mechanism specialized for mental states can be co-opted for use with nonmental states, presumably becoming more penetrable as it does so (Cohen et al., 2015). It is also the case for a specialization account involving an additional inhibitory selection process (e.g., Leslie et al., 2004).

However, the weakened effect size relative to previous research has ramifications for the confidence with which we can interpret these results in line with domain specialization. These results remain open to an alternative account by which adults are faster on false belief trials because they simply have more practice thinking about the content of beliefs than photographs. It is impossible to adjudicate between these two explanations without further data, but the key to resolving this issue will be to measure the size of any practice effect with beliefs and whether that would be sufficient to account for the small effect size found here. From these results, it would appear that there is no practice effect for *true* beliefs (over true photos or reality), which would appear to speak against the possibility of practice effects underlying the processing advantage for *false* beliefs. However, it is impossible to rule out the chance that participants were performing at close-to-maximum speed on all the "easier" trial types, meaning that practice effects would have no chance of being revealed with this design. Overall, we, therefore, prefer an interpretation by which our results as consistent with *either* specialization *or* greater experience of thinking about mental states than nonmental states such as the content of photographs.

A potential limitation to the present studies is that by contrasting false belief with false photos, we were contrasting not just representational formats but truth values; beliefs can be false because they can mislead, but it has been argued that so-called "false" photos can only be *outdated*, because they are judged to be accurate depictions of a past state (Leekam & Perner, 1991). In the present paper, we use the term "false photo" out of convention. This is a complicated issue with more than one possible consequence depending on whether one considers the processing of something false to be more difficult, the same as, or easier than processing something which is outdated. For the present finding to be overturned (i.e., for there to be no "true" difference between false photo and false belief processing), it would need to be demonstrated that outdatedness is *harder* than falseness in order to explain the slower response times on this trial type. We think this is highly unlikely because something which is false must exist in conflict with something true that is known to the participant. This conflict needs to be overcome for a correct response to be reached, and a processing cost should be expected. In contrast, something which is merely outdated does not conflict with what the participant knows to be true, so no such resolution is required. Instead, to our minds, it is more likely be that the higher cost of false photos relative to false beliefs here is *underestimated* if they were processed as outdated rather than false.

However, there is a further issue to consider. Commenting on an earlier draft of this manuscript, one reviewer argued that the false photos in our experiments might be false rather than outdated depending on the nature of the backstory. That is, since the photo was sent to a friend and meant to be informative (i.e., the present has been planted *here*), it might be considered misleading and, therefore, false. We think this is an interesting possibility, but we also feel that the question of what is misleading can be separated from the question of what is false, and that once this is done, the relative ease of false belief trials still requires explanation through another mechanism, be that specialization or practice effects. Specifically, we suggest that beliefs and photographs can be misleading if (and only if) their content would lead to a search error. Since the friend who received the photograph is not expected to search for the gift, the photo in this task cannot be misleading. However, it also follows that even the (uncontroversially) false belief of the woman who placed the gift in the drawer is not misleading; she has no reason to search for the item ever again as she planted it for someone else to discover. In short, neither false photos nor false beliefs in this task are misleading according to our definition. Therefore, since the higher cost for false photo trials over false belief trials is unlikely to be due to being outdated (for the reasons given above), and given they did not differ in "misleadingness," specialization or practice effects remain the most probable candidates to explain the cost of false photos.

Finally, these results also suggest that true beliefs are not processed more quickly than knowledge, contrary to the (unpredicted) results of some previous research (Cohen & German, 2009). The reason for the original finding could be due to the early use of the personal pronoun in sentences about beliefs. The absence of evidence for this hypothesis is not really a test of specialization as no theoretical model of belief processing has posited such an advantage. However, it is perhaps a limiting factor of any such specialization if it turns out to be limited to specifically *false* belief processing given the generally accepted view that the majority of beliefs that people hold are in fact true (Dennett, 1989; Leslie, 1994). Recall that true beliefs were not only no faster than reality trials but also no faster than true photo trials. An explanation for why specialization might be restricted to false mental states only is still required.

In conclusion, we investigated whether adults would be faster to process false beliefs than false photos, and true beliefs than reality. After removing the potential for variable influence of linguistic cues on different trial types and removing dual-task interference, we found evidence for a reliable advantage for false belief processing relative to false photos, albeit of a smaller magnitude than in previous research. These results are consistent with specialization accounts of mental state processing, but given the small size of the effect and the potential for the presumably relatively high frequency of belief processing in everyday life, it remains an open question as to whether these results support generalized practice effects instead.

Note

1 "who also contributed to the gift" was absent in Experiments 2 and 3.

Acknowledgments

We are very grateful to Anna Caunt for her assistance with data collection for Experiment 4.

Open Research Badges

This article has earned Open Data, Open Materials, and pre-registered badges. Data and materials are available at Exp 1: https://osf.io/prj67; Exp 2: https://osf.io/6fbpv; Exp 3: https://osf.io/cdk9a; Exp 4: https://osf.io/qszj2.

References

Apperly, I. A., Samson, D., Chiavarino, 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*(2), 300–321.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., & Bird, G. (2016). Can neurotypical individuals read autistic facial expressions? Atypical production of emotional facial expressions in autism spectrum disorders. *Autism Research*, *9*(2), 262–271.

Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition*, *111*(3), 356–363.

Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'. *Cognition*, *115*(3), 417–425.

Cohen, A. S., Sasaki, J. Y., & German, T. C. (2015). Specialized mechanisms for theory of mind: Are mental representations special because they are mental or because they are representations? *Cognition*, *136*, 49–63.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*, 568–570.

Dennett, D. C. (1989). *The intentional stance*. MIT Press.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, *6*(3), 346–359.

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.

Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, *40*(3), 203–218.

Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*(1–3), 211–238.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*(12), 528–533.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*(3), 225–251.

Mathot, S., Schreij, D., & Theeuwes, J. (2012). Open Sesame: An open-source, graphical experiment builder for the social sciences.

Milton, D. E. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, *27*(6), 883–887.

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

Samuel, S., Durdevic, K., Legg, E. W., Lurz, R., & Clayton, N. S. (2019). Is language required to represent others' mental states? Evidence from beliefs and other representations. *Cognitive Science*, *43*(1), 2313–2341.

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 thinking people: The role of the temporo-parietal junction in "theory of mind". *Neuroimage*, *19*(4), 1835–1842.

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728–755.

Williams, G. L. (2021). Theory of autistic mind: A renewed relevance theoretic perspective on so-called autistic pragmatic 'impairment'. *Journal of Pragmatics*, *180*, 121–130.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, *35*(1), 41–68.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting information 1
Supporting information 2
DataS1
DataS2
DataS3
DataS4