In Press with Psychological Science

(Association for Psychological Science, Blackwell Publishing)

Is belief reasoning automatic?

Ian A. Apperly[1], Kevin J. Riggs[2], Andrew Simpson[2], Claudia Chiavarino[1] and Dana Samson[1].

1. University of Birmingham, UK. 2. London Metropolitan University, UK.

Correspondence should be addressed to Ian Apperly:

School of Psychology
University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK
i.a.apperly@bham.ac.uk
+44 0121 414 3339

Word count 2473

Abstract

Understanding the operating characteristics of theory of mind is essential for understanding how beliefs, desires and other mental states are inferred, and the role such inferences could play in other cognitive processes. We present the first investigation of the automaticity of belief reasoning. In Condition 1 (incidental false belief task), adult subjects responded more slowly to unexpected questions concerning another person's belief about an object's location than to questions concerning the object's real location. Two further conditions showed that responses to belief questions were not necessarily slower than responses to reality questions, since subjects showed no difference in response times to belief and reality questions when they were instructed to track the person's beliefs about the object's location. The results suggest that adults do not ascribe beliefs to agents automatically.

Introduction

Reasoning about mental states such as beliefs, desires and intentions is the stock in trade of our everyday attempts to explain, predict and manipulate human behaviour. This ability – often termed "theory of mind" – is a fundamental component of human social cognition and of the uniquely human aptitude for communication (Baron-Cohen, Tager-Flusberg, & Cohen, 2001; Easton & Emery, 2005; Malle, Moses, & Baldwin, 2001; Repacholi & Slaughter, 2003; Sperber & Wilson, 1995). Surprisingly however, there have been few direct investigations of the basic operating characteristics of theory of mind. For example, it is unknown whether inferences about beliefs, desires and intentions are made automatically when we attend to the behaviour of agents, or whether such inferences are made ad hoc, according to need. Knowing whether theory of mind inferences are made automatically is critical for understanding how theory of mind interacts with other activities such as communication, and for improving the paradigms available for investigating theory of mind with event-related methods of cognitive psychology or neuroscience.

Several theorists argue that "theory of mind" processes such as belief reasoning must be automatic (Friedman & Leslie, 2004; Sperber & Wilson, 2002; Stone, Baron-Cohen, & Knight, 1998). This argument draws upon evidence that belief reasoning may depend upon cognitive processes that are domain-specific (Frith & Frith, 2003; Leslie & Thaiss, 1992; Saxe, Carey, & Kanwisher, 2004; though see Apperly, Samson, & Humphreys, 2005), or innate (Leslie, 2005; Onishi & Baillargeon, 2005). Domain-specificity and innateness are characteristic features of modular processes, and if theory of mind is modular, then it follows that processes such as belief reasoning may also be fast, informationally encapsulated and automatic (Fodor, 1983; Fodor, 2000; Friedman & Leslie, 2004; Leslie & Thaiss, 1992). The most detailed model of belief reasoning has been advanced by Leslie and colleagues (Friedman & Leslie, 2004; Leslie, German, & Polizzi, 2005; Leslie & Thaiss, 1992) who have argued that belief inferences are performed by a fast, automatic and domain specific theory of mind module that parses the behaviour of agents to generate a set of candidate belief contents. An executive control process is responsible for the second step of *selecting* a single belief content from among these candidates. No direct evidence bears upon the automaticity of either of these processing steps.

The current studies used a novel "incidental false belief task" to examine the automaticity of belief inferences. Our rationale was as follows. If subjects automatically parse events involving human agents in terms of the agents' beliefs, then making a later explicit judgement about those beliefs will depend upon information that has already been inferred and encoded. If this is the case then judgements about beliefs might be made as quickly as judgements about other information that was encoded about the event. Moreover, explicitly telling subjects to keep track of the agent's beliefs should not result in faster judgements about beliefs. In contrast, if subjects do not automatically infer and encode beliefs, then unexpected judgements about belief should be made relatively slowly compared with judgements about other information that *has* already been encoded. In this case, explicitly telling subjects to keep track of the agent's belief should result in faster judgements about belief because subjects would have the opportunity to infer the belief in advance. We therefore compared the speed of subjects' judgements to probe sentences about an agent's belief (where they thought an object was located) with the speed of judgements to probe sentences about reality (where the object was really located). Critically the speed of judgements to the same probe sentences was compared across conditions where we varied whether or not subjects were explicitly told to keep track of the agent's belief, or the corresponding reality.

Method.

The hypothesis that belief reasoning is automatic predicts that subjects should infer beliefs even in the absence of any particular reason to do so. *Condition 1: An Incidental False Belief Task* was devised so that subjects would monitor relevant aspects of reality but would have no particular reason to monitor agents' beliefs. We adapted video stimuli from Apperly, Samson, Chiavarino & Humphreys (2004) so that probe sentences could be presented at unpredictable intervals to elicit belief or reality judgements from subjects. In all trials a male actor hid an object in one of two boxes and a female actor indicated where she thought it was hidden. Subjects had to identify the location of the object at the end of each trial. To do so, he or she needed to monitor movement of the boxes and take account of whether the woman had a true or false belief when she gave her clue. Figure 1 depicts a generic event sequence for an experimental trial[1]. The subject saw the woman look in the boxes, then give her clue about the location of the object by placing a marker on one of them. By taking account of the fact that the woman's belief was true the subject could infer the object's location. The woman then left the room, and the man swapped the locations of the boxes. This had two effects. First, the subject needed to update their representation of the location of the object in order to solve the task of locating the object at the end of the trial. Second, the swap resulted in the woman having a false belief about the object's location. The change in the woman's belief state was not relevant to the task of locating the object at the end of the trial. Our interest was in whether subjects would, nonetheless, automatically infer the woman's new belief state. The video continued with the woman returning to the room. During this period the video paused and a probe sentence appeared: Either a belief probe "She thinks that it's in the box on the left" [or right on other trials] or a reality probe "It's true that it's in the box on the right" [or left on other trials]. Probes were presented approximately 3s, 6s or 9s after the boxes swapped and the woman's belief became false. After the subject responded to the probe the video continued until the appearance of a blue frame around the viewing area cued the subject to point to the location of the object. Since the only purpose of this component of the task was to encourage participants to keep track of information relevant for responding to reality probes, data from participants' pointing responses were not evaluated.

Since subjects needed to maintain and update a representation of the object's location in order to point correctly at the end of the trial we expected the reality probe to be answered with information that had already been processed (inferred and encoded). If subjects also maintained and updated a representation of the woman's belief, then belief probes would also be answered with information that had already been processed, and response times to belief probes might be no different from response times to reality probes. However, if subjects did not update their representation of the woman's belief automatically, then a correct response to the belief probe would require this extra information processing, which might result in a slower response time for belief probes than for reality probes.

Conditions 2 and 3 used the same video stimuli as Condition 1, and response times were recorded for the same probe sentences. The key difference from Condition 1 was that subjects in Conditions 2 and 3 were explicitly instructed to keep track of where the woman thought the object was located. Thus, subjects were expected to infer the woman's belief in advance of the belief probe, meaning that any processing cost associated with this inference would not be reflected in response times to belief probes. *Condition 2: Explicit belief and reality tracking*. Subjects were explicitly instructed to keep track of where the woman thought the object was located, where it was really located, and to point to the object's location at the end of each trial. *Condition 3: Explicit belief tracking.* Subjects were only instructed to keep track of where the woman thought the object was located. They were not asked to track where the object was located and were not required to point to the correct location of the object at the end of trials.

Insert Figure 1 near here

Subjects. Undergraduate students participated for course credits or a small honorarium: Condition 1, N=24 (15 Female, Mean age=21 years);

Condition 2, n=24 (14 Female, Mean age=20 years); Condition 3, N=26 (17 Female, Mean Age=21 years). Two subjects in condition 3 failed to complete the experiment and their data were not analysed.

Design and Procedure. There were 24 experimental trials, 12 with a belief probe and 12 with a reality probe. The correct answer for all experimental trials was “yes” and was equally likely to be left or right. A total of 56 filler trials were added to reduce the likelihood that subjects would be able to anticipate the exact event sequence in the video, or the timing, content or correct answer for the probe sentences. Twelve trials used the same videos and probes as the experimental trials, but the correct answer was “no”. Sixteen trials used the same videos as experimental trials, but the probes concerned physical facts other than the object’s location, or the knowledge state of the male actor. Sixteen trials combined all probe types with other videos using the same actors, objects and events, but in different sequences (from Apperly et al., 2004).

Altogether, 80 trials (each approximately 50s in length) were distributed over 4 experimental blocks, each comprising 6 experimental trials (3 belief probes, 3 reality probes) and a variety of filler probes. Over the block, correct answers were equally often “yes” and “no” and the object’s location at the time of the probe was equally often left or right. Trials were presented in a pseudo-random order, avoiding consecutive experimental trials. The experiment was presented on a standard Pentium-based desktop computer using DMDX (Forster & Forster, 2003). Response times were recorded from the onset of probe sentences, and so reflected the reading time for the probe sentences, plus any other processing required for responding “yes” or “no”.

Results

RTs falling 2 standard deviations beyond the mean per subject, per condition were removed. For belief probes this resulted in a loss of 10 data points (3.5%) in Condition 1, 11 data points (3.8%) in Condition 2 and 15 (5.2%) in Condition 3. For reality probes this resulted in a loss of 11 (3.8%) data points in Condition 1, 8 (2.7%) in Condition 2 and 10 (3.5%) in Condition 3.

Preliminary analyses showed no significant effect of the timing of the probes (3s, 6s or 9s after the woman’s belief becomes false), and although subjects responded more quickly with practice, this effect was similar for belief and reality probes. Data were collapsed across these factors for further analysis.

An ANOVA with probe type (belief, reality) as a within subject factor and Condition as a between-subject factor revealed a significant interaction between probe type and condition, $F(2,69)=3.49$, $p_{\text{rep}}=.93$, $\eta_p^2=.092$. The main effect of probe type failed to reach significance, $F(1,69)=3.45$, $p_{\text{rep}}=.90$ $\eta_p^2=.048$. The main effect of Condition was non-significant, $F(2,69)=.019$, $p_{\text{rep}}=.51$ $\eta_p^2=.001$. T-tests showed a significant difference between response times to belief and reality probes in Condition 1 (incidental false belief task), $t(23)=3.51$, $p_{\text{rep}}=.99$. The differences in Conditions 2 and 3 were non-significant: Condition 2 (explicit belief and reality tracking), $t(23)=.266$,

p_{rep}=.57; Condition 3 (explicit belief tracking), *t*(23)=.238, p_{rep}=.57. The mean response times are displayed in Figure 2.

Error analysis. Condition 1 belief probes, 35 (12%) incorrect; reality probes, 23 (8%) incorrect. This difference was not significant, but clearly indicates that the difference in response times to belief and reality probes did not result from a trade-off between speed and accuracy. Condition 2 belief probes, 36 (12.5%) incorrect; reality probes, 33 (11.5%) incorrect. Condition 3 belief probes, 23 (8%) incorrect; reality probes18 (6.3%) incorrect. Neither difference was significant.

Insert Figure 2 near here

Discussion

The incidental false belief task (Condition 1) showed a clear processing cost for belief probes in comparison with reality probes, consistent with subjects responding to reality probes using information they had already processed, but having to infer the woman's belief ad hoc in response to belief probes. This difference was absent in Conditions 2 and 3, suggesting that subjects could strategically infer the woman's belief in advance of the probes, and that having done so, belief probes were not intrinsically slower to process or respond to than reality probes. The fact that responses to reality probes were as fast as responses to belief probes in Condition 3 (in which subjects were only instructed to keep track of the woman's belief) suggests that it may not be possible to track the woman's false belief without also tracking the object's true location.

According to Leslie and colleagues (Friedman & Leslie, 2004; Leslie & Thaiss, 1992) a theory of mind module (ToMM) automatically parses the behaviour of agents to infer a set of candidate belief contents. However, the process of belief ascription is only complete once a separate, executive "selection-processor" *selects* the appropriate belief content for the agent from among these candidates. In these terms, responses to belief probes in the incidental false belief task could have been relatively slow because ToMM had not inferred candidate belief contents, or because subjects' "selection processor" had not yet selected the appropriate belief content from the set provided automatically by ToMM. In that latter case, the current results would still be compatible with the involvement of an automatic sub-process (such as ToMM) in belief ascription. But on either interpretation, the process of ascribing a belief to the woman is incomplete by the time the subject reaches the probe. Thus, whatever the nature of the sub-processes, the criterion for belief ascription – attributing a belief with a particular content to a particular individual – has not been met. In this most relevant sense, the current data suggest that belief reasoning is not automatic.

Our "incidental false belief task" addresses a significant methodological problem in the theory of mind literature, of knowing when a subject is making a theory of mind inference. The method identifies a narrow time window – immediately after the belief probe in the incidental false belief task – within which belief ascription should be taking place. Methods of this kind should enable event-related techniques from neuroscience and cognitive psychology to be used more effectively to investigate the nature of the complex component processes behind “theory of mind”.

Reference List

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and left temporo-parietal 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.

Apperly, I. A., Samson, D., & Humphreys, G. W. Domain specificity and theory of mind: Evaluating neuropsychological evidence. *Trends in Cognitive Sciences,* (in press).

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (2001). *Understanding other minds: Perspectives from developmental cognitive neuroscience*. (2 ed.) New York: OUP.

Easton, A. & Emery, N. J. (2005). *The cognitive neuroscience of social behaviour*. Hove: Psychology Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge MA: MIT press.

Fodor, J. A. (2000). *The mind doesn't work that way. The scope and limits of computational psychology.* Cambridge MA.: MIT Press.

Forster, K. L. & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavioral Research Methods, Instruments and Computers, 35,* 116-124.

Friedman, O. & Leslie, A. M. (2004). Mechanisms of belief-desire reasoning. Inhibition and bias. *Psychological Science, 15,* 547-552.

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

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences, 9,* 459-462.

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

Leslie, A. M. & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition: International Journal of Cognitive Science, 43,* 225-251.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). *Intentions and Intentionality: Foundations of social cognition*. Cambridge, MA: MIT Press.

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

Repacholi, B. & Slaughter, V. (2003). *Individual Differences in Theory of Mind - Implications for Typical and Atypical Development*. New York and Hove: Psychology Press.

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

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition*. (2 ed.) Oxford: Blackwell.

Sperber, D. & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language, 17(1&2),* 3-23.

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

Footnote

[1] Half of the experimental trials followed a similar sequence but the object was transferred from one box to the other in full view of the subject. Results did not differ for the two types of trial, and results are combined for all analyses.

Acknowledgements

This work was supported by a grant from the British Academy (SG-36063) to Ian Apperly.

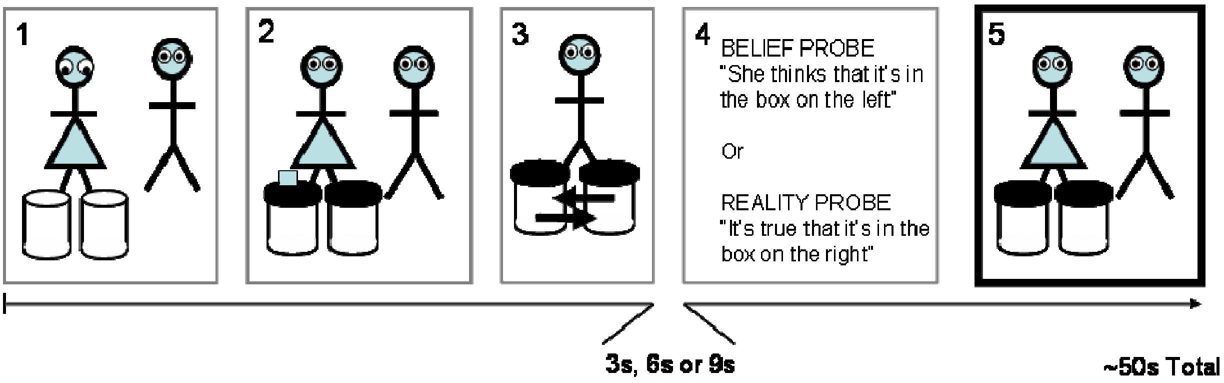
1
2
3
4
BELIEF PROBE
"She thinks that it's in the box on the left"
Or
REALITY PROBE
"It's true that it's in the box on the right"
5
3s, 6s or 9s
~50s Total

Figure 1. Schematic event sequence for experimental video trials with BELIEF / REALITY probes. 1. Woman looks in open boxes (so gains true belief about object's location). 2. Woman places marker to indicate location of object, then leaves room. 3. Man swaps boxes (so woman has false belief). 4. Probe sentence. 5. Woman returns and change in frame of video prompts subject to point to box containing object (Conditions 1 and 2 only)

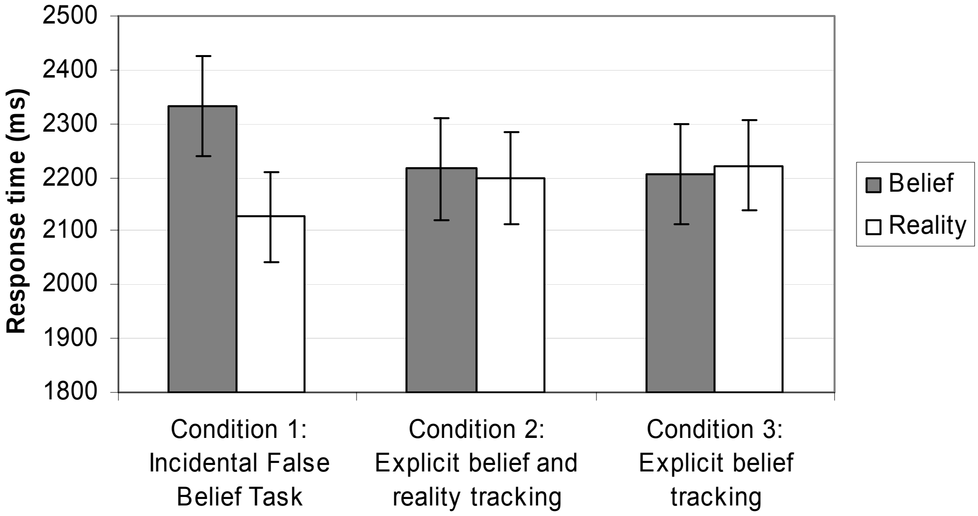

Figure 2. Mean response times (bars represent standard errors) for Belief and Reality probes in Conditions 1, 2 and 3.