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).