NeuroImage 61 (2012) 921–930

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Multiple roles for executive control in belief–desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning

Charlotte E. Hartwright *, Ian A. Apperly, Peter C. Hansen

University of Birmingham, UK

ARTICLE INFO

Article history:
Accepted 5 March 2012
Available online 14 March 2012

Keywords:
Theory of mind
Perspective taking
Executive control
Inhibition
Inferior frontal gyrus
Ventrolateral prefrontal cortex

ABSTRACT

Belief–desire reasoning is a core component of 'Theory of Mind' (ToM), which can be used to explain and predict the behaviour of agents. Neuroimaging studies reliably identify a network of brain regions comprising a 'standard' network for ToM, including temporoparietal junction and medial prefrontal cortex. Whilst considerable experimental evidence suggests that executive control (EC) may support a functioning ToM, co-ordination of neural systems for ToM and EC is poorly understood. We report here use of a novel task in which psychologically relevant ToM parameters (true versus false belief; approach versus avoidance desire) were manipulated orthogonally. The valence of these parameters not only modulated brain activity in the 'standard' ToM network but also in EC regions. Varying the valence of both beliefs and desires recruits anterior cingulate cortex, suggesting a shared inhibitory component associated with negatively valenced mental state concepts. Varying the valence of beliefs additionally draws on ventrolateral prefrontal cortex, reflecting the need to inhibit self perspective. These data provide the first evidence that separate functional and neural systems for EC may be recruited in the service of different aspects of ToM.

Introduction

The capacity to reason about the mental causes of action, termed 'mentalising' or exercising a 'Theory of Mind' (ToM), has received considerable interest from social neuroscientists over the last decade. Much attention has been given to identifying which, if any, brain regions should be considered as specialised for ToM. This work has made considerable progress in identifying possible contenders, and converges on the importance of a network of brain regions including temporoparietal junction (TPJ) and medial prefrontal cortex (mPFC) (Carrington and Bailey, 2009; Lieberman, 2007; Mar, 2011; van Overwalle, 2009). TPJ has been identified in the majority of neuroimaging studies of ToM, and appears selectively responsive when representing mental states such as beliefs, desires and intentions, over and above representation of physical states, personality traits or dispositions of the person, and above non-mental representations, such as photographs (Aichorn et al., 2009; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Saxe and Wexler, 2005). mPFC is also identified in most neuroimaging studies of ToM, though its activity may be less specific to mental state representation (Amodio and Frith, 2006), and may be most strongly recruited when reflecting on more enduring mental states, such as personality traits and social or moral beliefs (van Overwalle, 2009), or when making inferences under conditions of high uncertainty (Jenkins and Mitchell, 2010). The strong convergence of neuroimaging data has lead to a general consensus that TPJ and mPFC constitute the 'core' network for ToM, and that the functions they support are the most psychologically important for understanding ToM.

ToM has been studied most extensively using false belief tasks. A classic paradigm, the object transfer task, requires participants to make a prediction about the behaviour of a character, based upon the character's belief and desire at that point in time. A typical experimental sequence outlines a protagonist putting an object into location A. They then leave the scene. Whilst the protagonist is away, the object is transferred to location B. The character then returns, wishing to find the object but holding a false belief about its location (Wimmer and Perner, 1983). In order to successfully identify where the protagonist will look for the object, it is necessary for participants to infer the character's false belief about the object's location and predict the character's action on the basis of their false belief, whilst resisting interference from their own privileged knowledge of the object's true location, and what the right course of action would be. This task analysis leads to the expectation that successful ToM will not only require processes that might be specific to inferring and representing the mental states of others, but also processes for executive control (EC) to ensure that the correct information is selected for inferring mental states and predicting actions. It follows, then, that a complete account of the neural basis of ToM must also include brain regions associated with these sorts of control processes. To date, however, the brain bases of EC in ToM have been little explored.

* Corresponding author at: Centre for Behavioural Brain Sciences, 2.30 Hills Building, School of Psychology, University of Birmingham, B15 2TT, UK. Fax: +44 121 414 4897.
E-mail address: cee849@bham.ac.uk (C.E. Hartwright).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.03.012

Numerous researchers have noted that executive ability appears to contribute significantly to proficiency with ToM (e.g., Carlson and Moses, 2001; Carlson et al., 1998, 2002; Friedman and Leslie, 2004, 2005; German and Hehman, 2006; Leslie and Polizzi, 1998; Leslie et al., 2005; Perner and Lang, 1999; Wellman et al., 2001). For example, in the classic false belief paradigm mentioned above, children under the age of four seem unable to overcome their own knowledge of where the object really is. As a result, they consistently state that the protagonist will look for the object in the object's true location. Younger children, however, may sometimes pass the false belief task if the true location of the object is made less salient (Carlson et al., 1998; Wellman et al., 2001). These kind of 'egocentric errors', sometimes referred to as the 'curse of knowledge' (Birch and Bloom, 2004, 2007) or a 'reality bias' (Mitchell and Lacohee, 1991), have also been observed in older children and healthy adults (Bernstein et al., 2004; Birch and Bloom, 2007), and appear to reflect the need to exert EC to solve such tasks (see e.g., Apperly et al., 2005, 2009, for relevant discussions).

Two broad theoretical frameworks have been proposed concerning the role of EC in ToM. The first suggests that EC is necessary when a perspective difference between self and other exists, as is the case of false belief or conflicting desires. For example, knowledge of the true location may interfere with the ability to select the believed, or false, location. As a result, the self perspective must be inhibited in order to assume the perspective of the other (Ruby and Decety, 2003; Samson et al., 2005). This theory arises from behavioural observations of young children's propensity towards responding with their own knowledge, and data suggesting that performance in ToM tasks can be manipulated by varying the salience of self perspective (Carlson et al., 1998; Wellman et al., 2001). A growing literature suggests that the ventrolateral prefrontal cortex (vlPFC) may support this process of 'self perspective inhibition'. For example, Vogeley et al. (2001) identified that the right inferior frontal cortex, particularly right inferior frontal gyrus (rIFG), was modulated by varying the importance of self in a fictional scenario. This finding was later supported by a case study which demonstrated that damage to right vlPFC, including rIFG, resulted in interference from self perspective when attributing beliefs to others. In this particular case, the patient was able to solve ToM tasks where his own perspective was less salient, but failed ToM tasks where a clear incongruence between self and other knowledge state existed (Samson et al., 2005). Using false belief tasks from Samson et al.'s study and a stop-signal test of EC, a further study showed that the same ventral region of IFG was recruited bilaterally in healthy adults for both general response inhibition, and when contrasting false belief tasks that made high versus low demands on the inhibition of self-perspective (van der Meer et al., 2011). Finally, a recent study of visual perspective-taking showed an ERP component over right fronto-lateral cortex that was sensitive to differences between self and other perspectives (McCleery et al., 2011). These studies provide converging evidence that a functioning ToM is supported by regions outside of the 'standard' ToM network, and that the inferior frontal cortex – particularly vlPFC – may be an important, but overlooked, region involved in inhibition of self perspective. Notably, however, none of these studies examines the role of EC in reasoning about conative mental states, such as desires.

The second theory of the role of EC in ToM, proposed by Leslie and colleagues (Friedman and Leslie, 2004, 2005; Leslie and Polizzi, 1998; Leslie et al., 2005), extends beyond belief attribution to include the varying demands of desire reasoning. It is implicit in the standard false belief task that the character wishes to locate the object. However, if the agent holds a desire to avoid the object, both children and adults suffer further difficulty in false belief tasks (Apperly et al., 2011; Cassidy, 1998; German and Hehman, 2006). Moreover, like false belief reasoning, proficiency with avoidance desire coincides with the development of executive abilities. Leslie and colleagues explain this in terms of a shared inhibitory component for negatively valenced[1] mental states, such as false belief and avoidance desire. They suggest that, for both false belief and avoidance desire reasoning, participants are required to select from competing responses and inhibit the prepotent response (e.g., true versus believed location/desired versus undesired location). Consequently, false belief and avoidance desire states may draw on a domain-general 'selection processor', in order to direct executive selection resources in attentionally demanding situations. Importantly, avoidance desire (i.e. "desire to avoid") can concern objects or situations that are either intrinsically desirable or undesirable from the participant's own point of view. Hence, variation in the valence of desire does not reduce to a question of whether the participant shares the character's desire: indeed desire valence and self-other congruence of desire are logically orthogonal factors. To explain these findings, Leslie and colleagues suggest that EC has a more general role in ToM that is not restricted only to cases that require inhibition of self-perspective. Previous neuroimaging studies examining EC more generally in ToM have typically used separate tasks to identify ToM and EC regions. These indicate some overlap between neural regions recruited for EC tasks and false belief reasoning, extending beyond IFG to include anterior cingulate cortex (ACC), frontal operculum (FO) and frontal eyefields (FEF) (Rothmayr et al., 2011; Saxe et al., 2006b; van der Meer et al., 2011).

Whilst interest in the neural basis of executive function in ToM is growing, most previous studies are limited in their ability to cast light on the role of executive function in ToM. Most have sought to identify neural regions involved *only* in ToM by comparing activation observed in a ToM condition with that in a non-ToM control condition. Such approaches may enable powerful tests of hypotheses about brain regions that are domain-specific for ToM, but run the risk of subtracting out activation that is critical for understanding how ToM is achieved in the brain. A fruitful alternative approach is to manipulate psychologically relevant factors within a ToM task (for a discussion of these issues see discussion between Saxe et al., 2006a and Friston and Henson, 2006). Surprisingly few previous studies of ToM, however, have attempted such manipulations. Sommer et al. (2007) provide one of the few direct comparisons between true and false belief reasoning. In their nonverbal task, participants viewed a series of cartoons which depicted a true or false belief scenario analogous to the object transfer task outlined earlier. Regions which were more responsive to false belief over true belief attribution included the right TPJ, ACC and right lPFC. The reverse contrast only identified the superior frontal gyrus, which is in contention with the view that TPJ is an essential component when attributing any transient mental state (see van Overwalle, 2009). These data indicate that false belief reasoning might recruit EC regions. However, they are difficult to interpret with confidence, because it is not clear whether participants were solving the contrasting true belief condition by mental state ascription or by simply referring to the true state of affairs (Aichorn et al., 2009). Consequently, further examination of these two mental states is warranted, where attending to a protagonist's mental state is made unavoidable in both true and false belief reasoning. This was the case in the current study.

The neural basis of conative states such as desires has been studied less extensively. Hooker et al. (2008) examined neural activation when making empathic judgements for characters with varying perspectives. More directly relevant to the current study, Abraham et al. (2010) had participants read a series of short vignettes which varied the valence of belief and desire: either an agent's belief turned out to

[1] These variations in belief and desire both vary the difficulty of the belief–desire reasoning task. However, beliefs in the current study varied in terms of their consistency with the participant's self-perspective (true beliefs versus false beliefs), whereas desires varied only in terms of whether the target character liked or disliked the food. Therefore we use the term "valence" to refer collectively to these variations, so that true beliefs and desires for foods are described as "positively valenced" and false beliefs and desires to avoid foods are described as "negatively valenced".