Language and Linguistics Compass 3/1 (2009): 157–174, 10.1111/j.1749-818x.2008.00122.x

The Role of the Theory-of-Mind Cortical Network in the Comprehension of Narratives

Robert A. Mason* and Marcel Adam Just
Center for Cognitive Brain Imaging, Carnegie Mellon University

Abstract

Narrative comprehension rests on the ability to understand the intentions and perceptions of various agents in a story who interact with respect to some goal or problem. Reasoning about the state of mind of another person, real or fictional, has been referred to as Theory of Mind processing. While Theory of Mind processing was first postulated prior to the existence of neuroimaging research, fMRI studies make it possible to characterize this processing in some detail. We propose that narrative comprehension makes use of some of the neural substrate of Theory of Mind reasoning, evoking what is referred to as a *protagonist perspective* network. The main cortical components of this protagonist-based network are the dorsomedial prefrontal cortex and the right temporo-parietal junction. The article discusses how these two cortical centers interact in narrative comprehension but still play distinguishable roles, and how the interaction between the two centers is disrupted in individuals with autism.

Introduction

The psychological construct of Theory of Mind has been widely discussed in the field of developmental psychology and has influenced the fields of social psychology, philosophy, and neuroimaging. In its broadest sense, Theory of Mind refers to the ability to attribute internal mental states to others, as well as reasoning about one's own mental state. These attributed internal mental states can be intentions, feelings, beliefs, and emotions, among others (Baron-Cohen et al. 1985, Baron-Cohen 1988; Happé 1993; Tager-Flusberg 1993; Saxe, Carey and Kanwisher 2004). In the classic false belief task that has been used to study Theory of Mind, a child in a story is seeking an object that has been moved from its original location, without their knowledge. An observer who is able to reason about this situation using Theory of Mind will correctly predict that the child will look in the original location. Being able to understand the false belief of another person is one of the milestones in the development of Theory of Mind.

To understand the interactions of characters in a story, a reader has to attribute thoughts, goals, and intentions to the characters. It was not known,

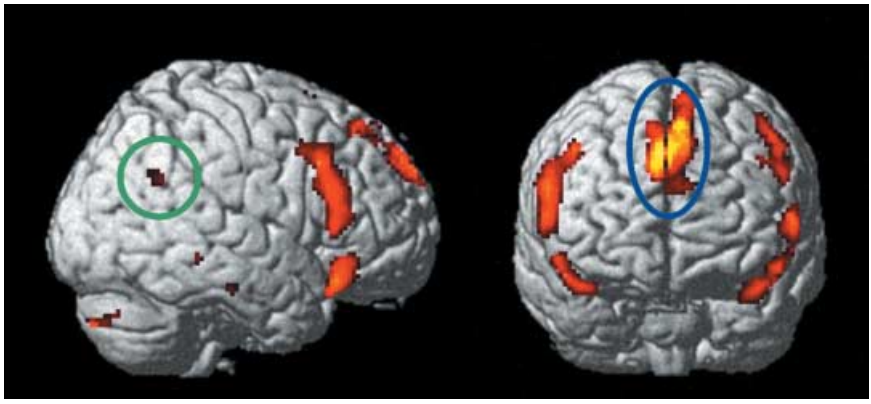

Fig. 1. Typical activation in the protagonist perspective network during a discourse task is depicted. The right temporo-parietal junction activation is highlighted by a green ellipse and the dorsomedial prefrontal cortex is indicated by a blue ellipse.

prior to the age of neuroimaging, that the considerations of others' mental states evoke a special cortical network. This network is not an inherent part of the linguistic processes in sentence comprehension; one does not need a Theory of Mind network to understand an asocial sentence like *The pen is on the desk*. But because of its role in the comprehension of narratives, the Theory of Mind network is part of the neural substrate of discourse comprehension. We have previously proposed that discourse comprehension involves a number of cortical networks, one of which is referred to as a protagonist (or agent) perspective interpreter network consisting of the dorsomedial prefrontal cortex (dmPFC) and the right temporal parietal junction (RTPJ), possibly extending into the posterior superior temporal sulcus (pSTS) as shown in Figure 1 (Mason and Just 2006). The term *protagonist* is used here for simplicity to refer to any human, animal, or other entity capable of autonomous action that is in focus in a story. Which character is focal may change throughout a narrative and there may be more than one focal character. In the remainder of the article, we explore the role of this protagonist perspective interpreter network (henceforth referred to as the protagonist perspective network) in the comprehension of narratives.

The cornerstone of our proposal is that Theory of Mind processing within narratives is best construed as the work of the protagonist perspective cortical network. One component of this network is a dorsomedial-prefrontal-based *protagonist monitor*, which can be viewed as an executive processor that activates throughout the processing of a narrative. The second network component, located near the right temporo-parietal junction, is a *protagonist simulator*, whose role may be to actively generate expectations of events based on an understanding of the intentions of the protagonist. Evidence supporting the characterization of this two-component protagonist perspective

network will be reviewed. An examination of the time course of the hemodynamic response will be offered as converging evidence for this characterization.

Neuroimaging research in narrative processing offers an unusual opportunity to examine the interaction between language comprehension and Theory of Mind processing in the understanding of protagonists' actions (see also, Gernsbacher et al. 1998; Mason et al. 2008). Consider for example this set of sentences:

> Brad had no money but he just had to have the beautiful ruby ring for his wife. Seeing no salespeople around, he quietly made his way closer to the ring on the counter. He was seen running out the door.

The passage invites the inference that Brad stole the ring. The inference is based on several items of information: Brad's desire and motivation to obtain the ring, his quiet manner in approaching a valuable object in a retail environment when no one was around, his impecuniousness, and his running out the door. Theory of Mind is necessary for interpreting the intentions, goals, and actions of characters within a narrative. The precise relation between such Theory of Mind processes that operate in narrative comprehension versus the ones that operate in traditional Theory of Mind tasks such as false belief (Wimmer and Perner 1983; Baron-Cohen et al. 1985; Gallagher and Frith 2003) remains to be determined. The following section describes the overlap in neural structures underlying the Theory of Mind processing in two types of tasks.

The remainder of the paper proceeds as follows. Section 2 describes the state of knowledge (and uncertainty) about the cortical regions implicated in Theory of Mind processing. Section 3 describes some evidence that suggests that more general processes underlie the Theory of Mind construct in narrative processing. Sections 4 and 5 provide some detail about the two centers that are proposed to constitute the protagonist perspective network; both centers also play roles in Theory of Mind processing, which is explained further in Section 6. Section 7 deals with the protagonist perspective network in language development and in autism, and Section 8 deals with the time course of activation in this network. Section 9 offers some brief conclusions.

The Neural Components of Theory of Mind

Neuroimaging studies have identified a network of regions supporting Theory of Mind processing, although there remains some debate concerning the roles of some of the regions. The initial neuroimaging studies of the Theory of Mind network focused primarily on non-discourse tasks. Many of these involved the presentation of cartoons, vignettes and animations (Fletcher et al. 1995; Brunet et al. 2000; Castelli et al. 2000, 2002; Galllagher et al. 2000; Martin and Weisberg 2003; Saxe and Kanwisher 2003; Schultz et al. 2003). These neuroimaging studies identified a set of

brain regions involved in Theory of Mind processing that includes the dorsomedial prefrontal cortex (dmPFC), the temporo-parietal junction (TPJ), the posterior superior temporal sulcus (pSTS), and the temporal poles (Fletcher et al. 1995; Brunet et al. 2000; Castelli et al. 2000, 2002; Gallagher et al. 2000; Vogeley et al. 2001).

There is a consensus that the frontal and posterior components of the Theory of Mind network have different roles, but there has been some debate concerning what these roles are. Gallagher and Frith (2003) proposed that the dorsomedial prefrontal cortex is the primary component of Theory of Mind reasoning, and that the right temporo-parietal junction simply provides the cues that are the input to the mentalizing (the reasoning about the mental state of another person). This is consistent with the finding that the dorsomedial prefrontal cortex activates during false belief stories, but not during true belief stories (Fletcher et al. 1995; Gallagher et al. 2000). By contrast, Saxe and Kanwisher (2003) proposed that the right TPJ is central to the attribution and interpretation of mental states. In contrast to Gallagher and Frith, Saxe and colleagues proposed that the role of dmPFC is not specific to Theory of Mind processing. Saxe and colleagues pointed out that although the dorsomedial frontal cortex may be involved in belief attribution, this region also activates during the processing of stories and vignettes that do not require false belief processing.[1]

A third possibility is that the roles of the frontal and posterior components change somewhat depending on the difficulty of the Theory of Mind reasoning that is required. For example, the frontal system alone may be adequate to keep track of protagonists' goals and intentions in very simple cases. However, when a text requires that the reader draw deep inferences regarding a character's mental state, the posterior part of the network may be additionally recruited. It may be possible to design text manipulations that determine whether the engagement of the right temporo-parietal junction region is simply a function of the difficulty of the text, or if the engagement is evoked by the need for a qualitatively different process. More generally, the operating characteristics of the Theory of Mind network and its component regions are open to fMRI experimentation. However, it is also possible that interactions of a network's components are so tightly bound to each other that disentangling their individual roles may be difficult using current methods.

Regardless of the precise computation that each area performs in Theory of Mind processing, the frontal and posterior regions appear to function in concert with each other. The main evidence for this coordination is provided by measures of the synchronization between the brain activity in the frontal and posterior centers in the Theory of Mind network, generally referred to as the measure of functional connectivity. The pairwise synchronizations among the centers in the Theory of Mind network cluster together (in a factor analysis, Koshino et al. 2005), indicating that the activity among these centers is coordinated. Moreover, in studies of people with

autism, who are known to have difficulty with Theory of Mind processing, the degree of synchronization among the components of the Theory of Mind network is lower than in a control group (Kana et al. 2008; Mason et al. 2008). We will return later to the insights about the Theory of Mind network provided by autism studies.

Reinterpreting Some of the Evidence for Theory of Mind Processing

A prominent neuropsychological study (Happé et al. 1999) showed that patients with right hemisphere damage (RHD) had a deficit in making a Theory of Mind-based causal inference in comparison to a non-mentalizing causal inference. One of the passages used by Happé et al. was:

> Tom dropped his glove while running from a store he had robbed. A police officer, who does not know that Tom is a burglar, sees the glove drop and calls Tom to alert him to the dropped glove.

The patients were then asked why the character did what s/he did.

Tompkins et al. (2008) proposed that some of the Happé evidence for Theory of Mind processing could be better construed as simply a RHD deficit at particularly demanding inference processing. Tompkins et al. argued that the control passages used by Happé et al. (1999) were not well matched to the Theory of Mind texts, placing the RHD patients at a disadvantage. In particular, the control passages did not contain an explicit text-based conflict and the Theory of Mind passages included more perspective representations than the control texts. Tompkins et al. showed that patients with RHD were not selectively impaired in reasoning about others' thoughts, beliefs and intentions when the materials were appropriately controlled for difficulty. One example of the better-matched physical inference passages was:

> Bob refuses to walk up the stairs to his 8th floor office. He always carries a heavy briefcase, and does not want to walk that far. One morning Bob, with his heavy load, is walking up the stairs along with many other people.

Additionally, Tompkins et al. replaced the query with a true–false probe. In this example, the probe was: *Bob saw that the elevator was working/leaving.* For the Theory of Mind causal inference passage, the probe was either related or unrelated to the character's belief, such as: *Tom thinks the officer knows he is innocent/studious.* The RHD patients in the Tompkins study showed no selective deficit on the mentalizing causal inference passages, calling into question a critical role of the right temporo-parietal junction. The RHD patients were equally impaired for the mentalizing and physical causal inference passages. This finding suggests that the deficit was more general and that it was the difficulty of the text that led to the impairment rather than the Theory of Mind processing. The Tompkins finding indicates that the role of the right temporo-parietal junction may not be specific to Theory

of Mind processing; instead, RTPJ may be involved in more general reasoning processes that are engaged during Theory of Mind tasks. A possible interpretation of the Tompkins et al. result is that this region becomes recruited only when an individual has to actively generate an expectation about a protagonist's state of mind as part of the comprehension of a difficult text. It becomes clear that the inconsistencies in these two studies bear on the degree to which strong conclusions about the functioning of the right hemisphere can be made from the neuropsychological literature alone.

Another perspective on the Theory of Mind network is that the functioning of the dorsomedial frontal cortex is not specific to Theory of Mind processing, but rather that it performs 'a domain-general initiation and maintenance of nonautomatic cognitive processes' (Ferstl and von Cramon 2002). Ferstl and von Cramon presented pairs of sentences that were either Theory-of-Mind-based (e.g., *Mary's exam was about to begin. Her palms were sweaty.*), or logic-based (e.g., *Sometimes a truck drives by the house. That's when the dishes start to rattle.*). As expected, when participants were given instructions to identify with the characters mentioned in the Theory of Mind passages, the dorsomedial prefrontal region was active in contrast to a control condition. The unexpected finding was that the dorsomedial prefrontal region was also active for the logic sentence pairs, even when participants were instructed to indicate whether there was a logical connection between the sentence pairs. This outcome provides a different perspective than earlier neuropsychological findings that had suggested that the dorsomedial prefrontal region was the primary center of Theory of Mind reasoning.

The implication of these two sets of results (Ferstl and von Cramon 2002; Tompkins et al. 2008) is that neither the dmPFC nor the RTPJ processes are specific to Theory of Mind *per se*. Instead, they are regions that activate during executive coherence monitoring (dmPFC) and reasoning during the reading of difficult texts (RTPJ).

Protagonist Monitor and the Dorsomedial Prefrontal Cortex

The common characterization of the dorsomedial prefrontal cortex as an executive processor suggests that the dorsomedial prefrontal cortex may flag and keep track of possible intention-related inferences. Thus, the activity of the dorsomedial prefrontal region of the protagonist perspective network may increase when information about the protagonist needs to be updated (particularly when an inference is required).

This characterization of the dorsomedial frontal function as an executive processor for discourse and not as a false belief processor is consistent with a case study which found no impairments in Theory of Mind processing in a patient with dorsomedial frontal lobe damage (Bird et al. 2004). However, this case study appears to be inconsistent with earlier neuropsychological evidence indicating that frontal lobe damage was associated with impairments in Theory of Mind processing (Stone et al. 1998; Channon and Crawford

2000; Rowe et al. 2001; Stuss et al. 2001). As was the case with the RHD patients, the neuropsychological data from frontal patients are not consistent enough to draw firm conclusions, due in part to the inability to assess how difficult a task is for a particular patient. Yet it is important to note that this literature can support the characterization that arises from a review of the fMRI literature of the dmPFC as a protagonist monitor.

Narrative-related activation has consistently been observed in the dorsomedial prefrontal cortex (Ferstl and von Cramon 2001, 2002; Ferstl et al. 2005; Xu et al. 2005; Kuperberg et al. 2006; Virtue et al. 2006; Mason et al. 2008). Xu et al. (2005, p. 1012) suggested that the dorsomedial prefrontal cortex operates 'at the interface between self and environment, yoking a variety of cognitive processes to knowledge about the world – a function that is clearly central to narrative comprehension'. In this view, processing discourse would be expected to engage systems that lie outside the language regions proper. An everyday understanding of others' minds is necessary for interpreting the intentions, goals, and actions of characters within a narrative. This same area has also shown activation in the comprehension of metaphor (Bottini et al. 1994) and identification of thematic roles within a story (Nichelli et al. 1995). All of these are discourse-level processes that are likely to be engaged during narrative comprehension.

The arguments concerning the role of dorsomedial prefrontal cortex in Theory of Mind processing have not converged. Frith and Frith (2003) tried to explain the Ferstl and von Cramon finding of activation in the dorsomedial prefrontal cortex in a task without Theory of Mind processing by suggesting that some of the stimulus materials may have invited attribution of beliefs and human actions, despite the absence of obvious people-related information in the stimuli or instructions. However, Saxe et al. (2004a) suggested the possibility that the 'coherence' instructions used by Ferstl and von Cramon in the 'logic' condition encouraged subjects to consider the intentions of the author of the passage. Alternatively, it is possible that, consistent with Ferstl and von Cramon's conclusion, the dorsomedial prefrontal cortex is involved in the more general process of protagonist monitoring. In the particular example given, the protagonist to be monitored was a non-human agent, namely, a truck. It is possible that the *truck* in the sample passage is treated and monitored as the protagonist.

Thus, the dorsomedial prefrontal area may be involved in more general protagonist-related processes, as indicated by its activation in narrative and Theory of Mind tasks. Although this brain area was found to be activated in many studies of discourse comprehension, it is not thought of as a language processing area. The dorsomedial prefrontal area had previously been linked to conceptual perspective-taking (e.g., Castelli et al. 2002). It had also been found to be active in tasks that required an understanding of the emotional and moral aspects of a situation (e.g., Greene et al. 2001). The recent discourse research has shown that this area also activates during the comprehension of narrative text and becomes activated in the processing

of character-centered emotional variables. However, activation of this region in non-discourse tasks (Krause et al. 1999) as well as 'inanimate' texts (Ferstl and von Cramon, 2002) suggests that the area has a domain-general role. In recognition of the domain-generality of this region's activation, Ferstl et al. (2007) designated the dorsomedial prefrontal cortex as part of the *Extended* Language Network (ELN). This label acknowledges that there is no sharp boundary between discourse processing and the processing of information about the world. In their meta-analysis of the neuroimaging of discourse literature, Ferstl and von Cramon reported that the dorsomedial prefrontal cortex was activated in several non-*Theory of Mind* tasks that focused on text coherence. While acknowledging that Theory of Mind might be a component part of any communication, they proposed that activation of the dorsomedial prefrontal cortex occurs in response to a 'domain-independent general process encompassing inferencing evaluation and Theory of Mind (Ferstl et al. 2007)'. The research demonstrating the activation of the dorsomedial prefrontal area in discourse processing shows just how informative neuroimaging research has been about discourse processing, but it also shows the limitations of the approach.

Protagonist Simulator and the Right Temporo-Parietal Junction (RTPJ)

Typically, Theory of Mind tasks activate portions of the right posterior, superior temporal gyrus and the right inferior parietal lobe. It is likely that these areas are part of the protagonist perspective network along with the previously discussed dorsomedial prefrontal area. Saxe and her colleagues (2004a, 2005) have proposed that the right temporo-parietal junction is critical for reasoning about others' mental states. Saxe and Wexler (2005) presented stories that varied across several attributes (familiar versus foreign background; normal behavior versus norm-violating behavior; character was satisfied versus character was not satisfied). Only the RTPJ (but not dmPFC or LTPJ) consistently showed activation in conditions associated with mentalizing about the protagonist. The time course of the hemodynamic response in the RTPJ revealed an increase in activation at the time when the mental state of the protagonist was described; there was no similar modulation of the hemodynamic in response to a description of background information. Moreover, there was an additional increase in the RTPJ response when the protagonist's background and mental state were incongruent. These findings establish a role for RTPJ when a reader must consider the reasons for the behavior of the protagonist.

Activation has also been observed in the temporal poles adjacent to the amygdala in Theory of Mind tasks (Fletcher et al. 1995; Gallagher et al. 2000; Brunet et al. 2000; Castelli et al. 2000, 2002; Vogeley et al. 2001), but the activation may not be related to Theory of Mind processing. This area consistently activates during text comprehension regardless of whether or not there is a false or true belief present. In Ferstl's meta-analysis of text

comprehension neuroimaging, the anterior temporal lobe was consistently found to be activated in many of the contrasts (e.g., as a function of coherence, language, rest, and text interpretation processes). Ferstl et al. proposed that these anterior temporal lobe areas are involved in integrating levels of language into a text representation. This is consistent with the proposal that the anterior temporal regions are part of a text integration network (Mason and Just 2006).

Also worth noting is that the posterior superior temporal activation that has been found during the viewing of simple interactions of moving objects (Blakemore and Decety 2001, Blakemore et al. 2003; Schultz et al. 2004; Pelphrey et al. 2005) and need not be associated with attribution of mental states, as Saxe et al. (2004a,b) have pointed out. Saxe and colleagues attributed early developing social concepts (e.g., perception, desire, and emotion) to the pSTS rather than to the mentalizing role of the more posterior and superior temporo-parietal junction. This proposal is consistent with the right superior temporal sulcus being a part of Ferstl's ELN network (2007). They found that the pSTS was part of a coherence network. According to this view, the posterior superior temporal sulcus is not critically involved in mentalizing.

The Interaction of the Protagonist Perspective Network Components

We propose that the dorsomedial prefrontal cortex and the right temporo-parietal junction perform different, but nevertheless highly interdependent functions during discourse comprehension. The dorsomedial prefrontal cortex may flag and keep track of possible intention-related inferences, whereas the right temporo-parietal junction may be more involved with reasoning about particular mental states (Saxe and Kanwisher 2003). These proposed functions are consistent with the general view of the dorsomedial prefrontal cortex as an executive processor and with the description of this region's putative role in Theory of Mind tasks (Gallagher and Frith 2003). Thus, the dorsomedial prefrontal region of the protagonist perspective network may become more activated when information about the protagonist needs to be updated (particularly when an inference is required), whereas the posterior portion of the network may become activated only when the protagonist's intention needs to be processed. It is also possible that the tempo-parietal activation for intentional inferences could be related to reasoning about a protagonist's intention based on one's own experience.

The dorsomedial prefrontal cortex seems to be particularly well-suited to processing information related to understanding another person's plans and motives. This can also be viewed as comprehension of an alternative reality, specifically the world-view of a protagonist within a text. Any inference that is related to a characteristic of a protagonist should result in activity in this region as part of the updating of the protagonist perspective. The dorsomedial prefrontal region has also shown increased activity when

a text is emotionally oriented or requires the reader to reference emotionally based memories (Ferstl et al. 2005). Activation in this dorsomedial prefrontal area has also been observed in tasks that involve emotional processing, and more specifically, emotion related to memory (Canli et al. 2002; Nakic and Gabrielli, unpublished manuscript). Thus, the activation in dorsomedial prefrontal areas might be attributed in part to emotion-related processing. Dorsomedial prefrontal activation has also been observed in tasks that require judgments concerning self-referential episodic memories (Zysset et al. 2002). The fact that the dorsomedial prefrontal region is often activated in discourse tasks, social cognitive processing, and Theory of Mind suggests that this cortical area plays a general role that may be common to many of these types of processes.

Despite our proposal that RTPJ and dmPFC have particular roles in discourse comprehension, we also propose that these regions (and perhaps most cortical regions) can support several different types of functions. The multiple functions of a region presumably have in common an underlying computational style stemming ultimately from the operating characteristics of the neurons in the region. As a given task is encountered, a particular region is likely to be apart of the evoked cortical network if one of its specialized computational capabilities is called for by the task. In particular, the cortical constituents of the protagonist network co-activate if the task (such as reading a narrative) and a reader's goals (comprehension) entail understanding the thoughts and intentions of a person participating in a sequence of events. This formulation of cognitive roles for these two regions is not meant to be an exclusive localization of function. Mitchell (2008) has recently criticized the conceptualization of right temporo-parietal junction and the dorsomedial prefrontal cortex as Theory of Mind regions by demonstrating that right temporo-parietal junction participates in other types of processing (RTPJ was modulated by a nonsocial attention task). Corbetta and Shulman (2002) have made a similar argument. Interestingly, Mitchell proposes that a 'suggestive – and disquieting – account of the current data is that the RTPJ subserves entirely different processes as a function of the other brain regions engaged by a particular task'. We do not find this disquieting at all, but a sensible way for versatile neural computational systems to be deployed.

Time Course of Activation in the Protagonist Network

The time courses of the activation of the protagonist network's component nodes provide useful information about their roles. Saxe and colleagues (2004a, 2006) as well as other researchers have examined the activation in a functional region of interest (fROI), in conjunction with whole brain voxel analysis, to investigate the neural basis of Theory of Mind. This focus on the time course is particularly salient in discourse comprehension research, which involves an inherent distribution of the processing over many seconds.

Although the temporal resolution of fMRI is inferior to that of ERPs (see van Burkum (2004) for the use of ERP to study discourse processing), the *relative* fMRI time courses of the various areas provide insights into the relationship between regions.

Mason et al. (2008) examined the time course of the activation within key regions (such as dmPFC and RTPJ) and the functional connectivity between them, illuminating the relation among the components of the protagonist perspective network. Participants read passages that invited inferences concerning a character's intentions (the earlier passage about Brad stealing a ring provides an example), as well as passages that invited an inference based on physical causality. An example of a physical causality passage is:

> While playing in the waves, Sarah's Frisbee went flying toward the rocks in the shallow water. While searching for it, she stepped on a piece of glass. Sarah had to wear a bandage on her foot for a week.

Interestingly, the activity in the dorsomedial prefrontal cortex was similar across the physical and intentional conditions, and it was also similar in the context-setting sentences at the beginning of the passages. This finding suggests that the participants were monitoring a protagonist throughout the text. In contrast, there was an increase in the activation of the right temporo-parietal junction only when an inference was required. RTPJ became activated as soon as it was possible for the reader to actively simulate a protagonist's intention. In the sample passages, this can occur as early as the second sentence. This type of predictive inference has also been seen in other neuroimaging studies of inference-making (Virtue et al. 2006).

The time course of activation differed between the two components of the Theory-of-Mind-based protagonist network. The dorsomedial prefrontal region activated early in the reading of both types of narratives (involving either a physical or an intentional inference), probably in response to the need to monitor a protagonist or an agent, as shown in Figure 2A. It remains to be seen what types of protagonist-related information particularly evoke activation in this region. In contrast, the right temporo-parietal junction showed a marked increase in activity when an inference was invited, as shown in Figure 2B. Critically, the RTPJ was the only inference-related activating region in which the signal intensity increase was greater for the intentional inferences than the physical inferences. This intention-based mentalizing activation is consistent with the view of RTPJ supporting mental state reasoning (Saxe et al. 2004a). However, the findings do not rule out the possibility that the role of RTPJ is to collate the cues to be used as input to mentalizing (Gallagher and Frith 2003).

The Protagonist Perspective Network in Development and Autism

In the study of language development, a key issue has concerned the time at which Theory of Mind processing emerges relative to the emergence of

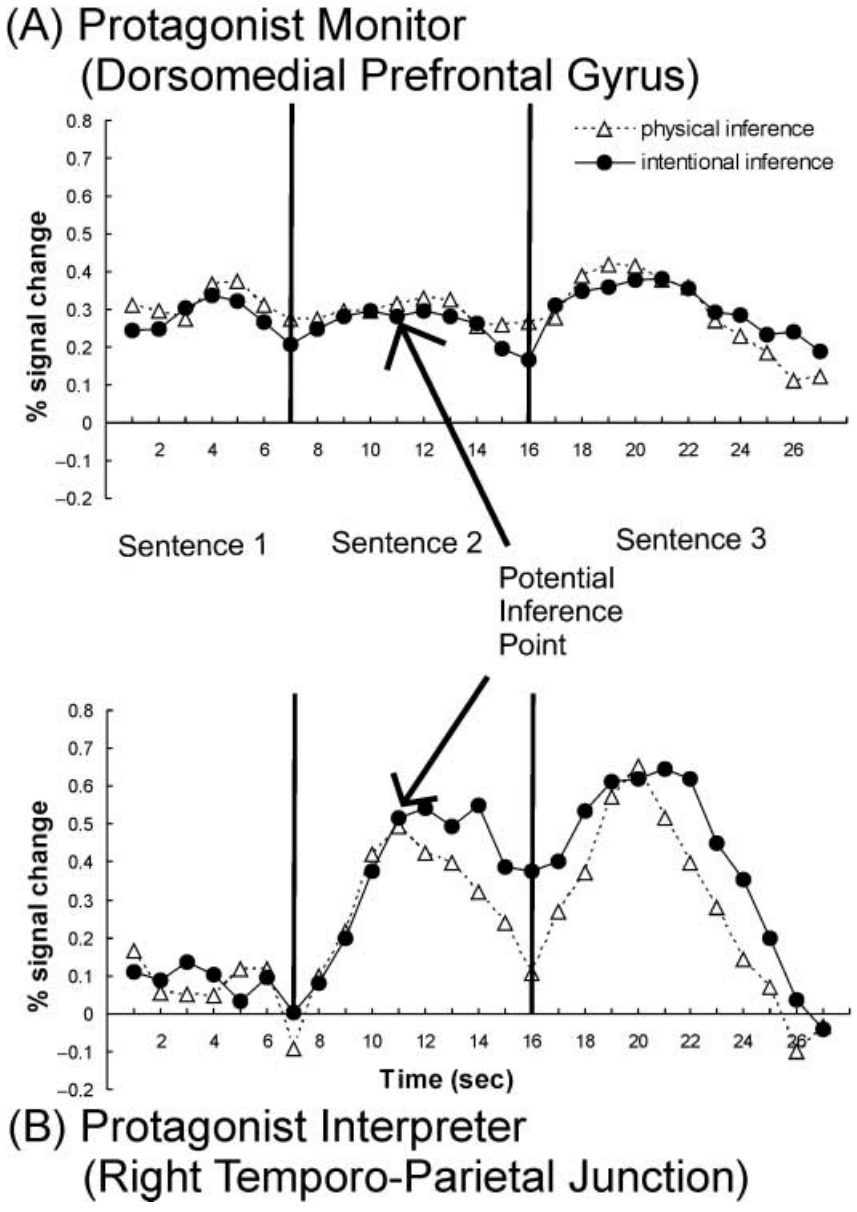

Fig. 2. The average time course of activated voxels in the two functional regions of interest (ROIs) during the reading of passages inviting intentional inferences and of those inviting physical inferences indicate that (A) the activation for the two inference types is similar across sentences for the protagonist monitor (dorsomedial prefrontal cortex) and (B) the activation for the protagonist simulator (right temporo-parietal junction) is different for the two inference types, diverging at the earliest measurable peak activation of a potential inference. The activation following the divergence is greater during the reading of intentional inferences than physical inferences.

various language processes (see Malle 2002 for an overview). Theory of Mind develops over several years with increasing complexity, emerging at around the age of 2 for desires, at around age 3 for beliefs, and at around age 4 for the understanding of false beliefs (Wellman 1990). The first Theory of Mind processing coincides approximately with the first growth spurt in vocabulary development (18–20 months), whereas the 2- to 4-year interval coincides with the acquisition of syntactic knowledge. Language development and Theory of Mind development influence each other. For example, knowledge of the speaker's intentions (Theory of Mind) can influence learning of novel words (Sabbagh and Baldwin 2001). Conversely, understanding of sentences with embedded complements (language development) can predict children's mastery of false belief (de Villiers 2000). The co-development in the two abilities suggests that there may be similar or interacting neural components for language processing and theory of mind.

The neural systems of individuals with autism, who have deficits in Theory of Mind processing, provide further insights into the relationship between language and Theory of Mind. Individuals with autism have exhibited deficits in Theory of Mind processing in many behavioral studies (e.g., Baron-Cohen et al. 1985, 1995; Perner et al. 1989; Reed and Peterson 1990; Leekam and Perner 1991; Swettenham 1996; Swettenham et al. 1996), as well as in several neuroimaging studies (e.g., Happé et al. 1996; Castelli et al. 2002; Schultz et al. 2003). More recently, neuroimaging evidence concerning the protagonist monitoring network in autism has become available. These recent neuroimaging data indicate how people with high-functioning autism deal with the challenge of understanding the actions and intentions of others as part of discourse comprehension. An additional part of the challenge is the need to infer and integrate a myriad of facts about the world that have to be inferred and integrated during discourse comprehension.

Mason et al. (2008) proposed that the cortical system in autism attempts to meet the challenges by engaging the right hemisphere Theory of Mind areas during narrative comprehension, regardless of whether they are called for. In the autism group, there was greater activation (relative to controls) in the right temporo-parietal junction and the dorsomedial prefrontal areas in all of the experimental conditions, regardless of whether the text entailed a Theory of Mind inference. By contrast, the control group showed increased activation in these areas only in the Theory of Mind conditions (which entailed an inference about a protagonist's intention). The unselective Theory of Mind processing in the participants with autism may reflect a compensatory Theory of Mind vigilance, always on guard for possible Theory of Mind information. Thus the challenge of accomplishing protagonist monitoring in autism in the face of a damaged Theory of Mind cortical network appears to be met by unselectively activating the network during the reading of all texts.

An additional telling characteristic of the Theory of Mind network in autism emerged. The functional connectivity (synchronization of activation) among the nodes of the Theory of Mind network was lower than in controls, indicating an inefficiency or disruption of communication among the nodes of this network in autism. The frontal-posterior underconnectivity in autism may limit how effectively the recruited cortical areas can function as an integrated network. This frontal-posterior functional underconnectivity in autism affects not only the Theory of Mind network, but any cortical network with a frontal component, as demonstrated by findings of frontal posterior underconnectivity in tasks such as working memory for faces (Koshino et al. 2005), executive function (Just et al. 2007), comprehension of highly visual sentences (Kana et al. 2006) and high-level inhibition (Kana et al. 2007). The disruption of the Theory of Mind network in autism in discourse processing is emblematic of the more general difficulty in autism of making sense of the events of everyday life. More generally, the autism findings point out the centrality of Theory of Mind processing in everyday life.

Conclusions

The ability to understand the world from the perspective of a protagonist is a critical component in comprehending a narrative. During such comprehension, readers often generate expectations about the actions of protagonists based on an understanding of the character's intentions. We have proposed that this Theory of Mind processing is supported by a protagonist perspective cortical network. The first component of this network is a dorsomedial-prefrontal-based protagonist monitor, which can be viewed as an executive processor that activates throughout the processing of a narrative. Future research is necessary to examine what types of information about the protagonist enter into the computations and affect the activation of this component of the network. The second network component, located near the right temporo-parietal junction, is a protagonist simulator, whose role may be to actively simulate expectations based on an understanding of the intentions of the protagonist. This information can then be integrated into a reader's situation model (or mental model) of the text. Recent results in functional connectivity in individuals with autism, as well as an examination of the time course of activation within functional regions of interest, provide initial support for these functional distinctions, which can generate hypotheses for future research on Theory of Mind processing in discourse comprehension.

Notes

* Correspondence: Robert Mason, Center for Cognitive Brain Imaging, Department of Psychology, Carnegie Mellon University. Pittsburgh, PA 15213, USA. Tel: (412) 268-3784; Fax: (412) 268-2804. E-mail: rmason@andrew.cmu.edu. This research was supported by the

National Institute of Mental Health Grant MH029617 and Autisim Centers of Excellence Grant HD055748 from the National Institute of Child Health and Human Development

[1] Saxe et al. (2004a) went on to distinguish between two posterior areas involved in Theory of Mind processing, namely the right posterior superior temporal sulcus and the right temporo-parietal junction. In their proposal, it was the right temporo-parietal junction that was the center of mentalizing, whereas the more anterior region (the posterior superior temporal sulcus region) was associated with violations of expectations about human behavior during perception (Pelphrey 2006) and in the representation of actions (Decety et al. 2002).

References

Baron-Cohen, S. 1988. Social and pragmatic deficits in autism: cognitive or affective? Journal of Autism and Developmental Disorders 18.379–402.

——. 1995. Mindblindness: An essay on autism and theory of mind. Cambridge, MA: MIT Press.

——, A. M. Leslie, and U. Frith. 1985. Does the autistic child have a 'theory of mind'? Cognition 21.37–46.

Blakemore, S. J. and J. Decety. 2001. From the perception of action to the understanding of intention. Nature Reviews Neuroscience 2.561–7.

——, P. Boyer, M. Pachot-Clouard, A. Meltzoff, C. Segebarth, and J. Decety. 2003. The detection of contingency and animacy from simple animations in the human brain. Cerebral Cortex 13.837–44.

Bottini, G., R. Corcoran, R. Sterzi, E. Paulesu, P. Schenone, P. Scarpa, R. S. J. Frackowiak, and C. D. Frith. 1994. The role of the right hemisphere in the interpretation of figurative aspects of language: a positron emission tomography activation study. Brain 117.1241–53.

Brunet, E., Y. Sarfati, M. C. Hardy-Bayle, and J. Decety. 2000. A PET investigation of the attribution of intentions with a nonverbal task. NeuroImage 11.157–66.

Canli T., J. E. Desmond, Z. Zhao, and J. D. Gabrieli. 2002. Sex differences in the neural basis of emotional memories. Proceedings of the National Academy of Sciences 99.10789–94.

Castelli, F., C. Frith, F. Happé, and U. Frith. 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain 125.1839–49.

——, F. Happé, U. Frith, and C. D. Frith. 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. NeuroImage 12.314–25.

Channon, S. and S. Crawford. 2000. The effects of anterior lesions on performance on a story comprehension test: left anterior impairment on a theory of mind-type task. Neuropsychologia 38.1006–17.

Corbetta M. and G. L. Shulman. 2002. Control of goal-directed and stimulus driven attention in the brain. Nature Reviews Neuroscience 3.201–15.

De Villiers, J. (2000). Language and theory of mind: what are the developmental relationships? Understanding other minds: Perspectives from developmental cognitive neuroscience, ed. by S. Baron-Cohen, H. Tager-Flusberg and D. J. Cohen, 2nd edn, 83–123. New York: Oxford University Press.

Decety, J., T. Chaminade, J. Grèzes, and A. N. Meltzoff. 2002. A PET exploration of the neural mechanisms involved in reciprocal imitation. NeuroImage 15.265–72.

Ferstl, E. C. and D. Y. von Cramon. 2001. The role of coherence and cohesion in text comprehension: an event-related fMRI study. Cognitive Brain Research 11.325–40.

——. 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? NeuroImage 17.1599–612.

Ferstl, E. C., M. Rinck, and D. Y. von Cramon. 2005. Emotional and temporal aspects of situation model processing during text comprehension: an event-related fMRI study. Journal of Cognitive Neuroscience 17.724–9.

——, J. Neumann, C. Bogler, and D. Y. von Cramon. 2007. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. Human Brain Mapping 29.581–93.

Fletcher, P. C., F. Happé, U. Frith, S. C. Baker, R. J. Dolan, R. S. J. Frackowiak, and C. D. Frith. 1995. Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. Cognition 57.109–28.

Frith, U. and C. D. Frith. 2003. Development and neurophysiology of mentalizing. Philosophical Transactions of the Royal Society of London Series B – Biological Sciences 358.459–73.

Gallagher, H. L. and C. D. Frith. 2003. Functional imaging of 'Theory of Mind'. Trends in Cognitive Science 7.77–83.

Gallagher, H. L., F. Happé, N. Brunswick, P. C. Fletcher, U. Frith, and C. D. Frith. 2000. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia 38.11–21.

Gernsbacher, M. A., B. M. Hallada, and R. R. W. Robertson. 1998. How automatically do readers infer fictional characters' emotional states? Scientific Studies of Reading 2.271–300.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293.2105–8.

Happé, F. G. 1993. Communicative competence and theory of mind: a test of relevance theory. Cognition 48.101–19.

Happé, F., S. Ehlers, P. Fletcher, U. Frith, M. Johansson, C. Gillberg, R. Dolan, R. Frackowiak, and C. Frith. 1996. 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. Neuroreport 8.197–201.

Happé, F. G., H. Brownell, and E. Winner. 1999. Acquired 'Theory of Mind' impairments following stroke. Cognition 70.211–40.

Just, M. A., V. L. Cherkassky, T. A. Keller, R. K. Kana, and N. J. Minshew. 2007. Functional and anatomical cortical underconnectivity in autism: evidence from an fMRI study of an executive function task and corpus callosum morphometry. Cerebral Cortex 17.951–61.

Kana, R. K., T. A. Keller, V. L. Cherkassky, N. J. Minshew, and M. A. Just. 2006. Sentence comprehension in autism: thinking in pictures with decreased functional connectivity. Brain 129.2484–93.

Kana, R. K., T. A. Keller, N. J. Minshew, and M. A. Just. 2007. Inhibitory control in high functioning autism: decreased activation and underconnectivity in inhibition networks. Biological Psychiatry, 62.198–206.

Kana, R. K., T. A. Keller, V. L. Cherkassky, N. J. Minshew, and M. A. Just. 2008. Atypical frontal-posterior synchronization of Theory of Mind regions in autism during mental state attribution. Social Neuroscience 99999.1–18.

Koshino, H., P. A. Carpenter, N. J. Minshew, V. L. Cherkassky, T. A. Keller, and M. A. Just. 2005. Functional connectivity in an fMRI working memory task in high-functioning autism. NeuroImage 24.810–21.

Krause, B. J., D. Schmidt, F. M. Mottaghy, J. Taylor, U. Halsband, H. Herzog, L. Tellmann, and H. Muller-Gartner. 1999. Episodic retrieval activates the precuneus irrespective of the imagery content of word pair associates: a PET-study. Brain 122.255–63.

Kuperberg, G. R., B. M. Lakshmanan, D. N. Caplan, and P. J. Holcomb. 2006. Making sense of discourse: an fMRI study of causal inferencing across sentences. NeuroImage 33.343–61.

Leekam, S. and J. Perner. 1991. Does the autistic child have a metarepresentational deficit? Cognition 40.203–18.

Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. The evolution of language out of pre-language, ed. by T. Givón and B. F. Malle, 265–84. Amsterdam: Benjamins.

Martin, A. and J. Weisberg. 2003. Neural foundations for understanding social and mechanical concepts. Cognitive Neuropsychology Special Issue: the Organization of Conceptual Knowledge in the Brain: Neuropsychological and Neuroimaging Perspectives 20.575–87.

Mason, R. A. and M. A. Just. 2006. Neuroimaging contributions to the understanding of discourse processes. Handbook of psycholinguistics, ed. by M. Traxler and M. A. Gernsbacher, 765–99. Amsterdam: Elsevier.

Mason, R. A., D. L. Williams, R. K. Kana, N. Minshew, and M. A. Just. 2008. Theory of mind disruption and recruitment of the right hemisphere during narrative comprehension in autism. Neuropsychologia 46.269–80.

Mitchell, J. P. 2008. Activity in the right temporo-parietal junction is not selective for theory-of-mind. Cerebral Cortex 18.262–71.

Nichelli, P., J. Grafman, P. Pietrini, K. Clark, K. Y. Lee, and R. Miletich. 1995. Where the brain appreciates the moral of a story. Neuroreport 6.2309–13.

Pelphrey, K. A., and J. P. Morris. 2006. Brain mechanisims for interpreting the actions of others from biological-motion cues. Current Directions in Psychological Science 15.136–40.

Pelphrey, K. A., J. P. Morris, C. R. Michelich, T. Allison, and G. McCarthy. 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. Cerebral Cortex 15.1866–76.

Perner, J., U. Frith A. M. Leslie and S. Leekam. 1989. Exploration of the autistic child's theory of mind: knowledge, belief, and communication. Child Development 60.689–700.

Reed, T. and C. Peterson. 1990. A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. Journal of Autism and Developmental Disorders 20.555–68.

Rowe, A. D., P. R. Bullock, C. E. Polkey, and R. G. Morris. 2001. 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. Brain 124.600–16.

Sabbagh, M. A. and D. A. Baldwin. 2001. Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. Child Development 72.1054–70.

Saxe, R. 2006. Why and how to study Theory of Mind with fMRI. Brain Research, 1079.57–65.

Saxe, R. and N. Kanwisher. 2003. People thinking about thinking people. The role of the temporo-parietal junction in 'Theory of Mind'. NeuroImage 19.1835–42.

Saxe, R. and A. Wexler. 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43.1391–9.

Saxe, R., S. Carey, and N. Kanwisher. 2004a. Understanding other minds: linking developmental psychology and functional neuroimaging. Annual Review of Psychology 55.87–124.

Saxe, R., D.-K. Xiao, G. Kovacs, D. I. Perrett, and N. Kanwisher. 2004b. A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42.1435–46.

Schultz, R. T., D. J. Grelotti, A. Klin, J. Kleinman, C. Van der Gaag, R. Marois, and P. Skudlarski. 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. Philosophical Transactions of the Royal Society of London Series B – Biological Sciences 358.415–27.

Schultz, J., H. Imamizu, M. Kawato, and C. D. Frith. 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. Journal of Cognitive Neuroscience 16.1695–705.

Stone, V. E., S. Baron-Cohen, and R. T. Knight. 1998. Frontal lobe contributions to theory of mind. Journal of Cognitive Neuroscience 10.640–56.

Stuss, D. T., G. G. Gallup, and M. P. Alexander. 2001. The frontal lobes are necessary for theory of mind. Brain 124.279–86.

Swettenham, J. 1996. Can children with autism be taught to understand false belief using computers? Journal of Child Psychology and Psychiatry 37.157–65.

Swettenham, J., S. Baron-Cohen, J. C. Gomez, and S. Walsh. 1996. What's inside a person's head? Conceiving of the mind as a camera helps children with autism develop an alternative theory of mind. Cognitive Neuropsychiatry 1.73–88.

——. 1993. What language reveals about the understanding of minds in children with autism. Understanding other minds: Perspectives from autism, ed. by S. Baron-Cohen, H. Tager-Flusberg, and D. J. Cohen, 138–57. Oxford, UK: Oxford University Press.

——. 1997. Language acquisition and theory of mind: contributions from the study of autism. Research on communication and language disorders: Contributions to theories of language development, ed. by L. B. Adamson and M. A. Romski, 135–60. Baltimore, MD: Paul Brookes Publishing.

Tompkins, C. A., V. L. Scharp, W. Fassbinder, K. Meigh, and E. M. Armstrong. 2008. 'Theory of Mind' in adults with right hemisphere brain damage. Aphasiology 22.42–61.

Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: can we use ERPs to keep track of things? The on-line study of sentence comprehension: Eyetracking, ERPs and beyond, ed. by M. Carreiras and C. Clifton Jr. 229–70. New York: Psychology Press.
Virtue, S., J. Haberman, Z. Clancy, T. Parrish, and M. Jung-Beeman. 2006. Neural activity of inferences during story comprehension. Brain Research 1084.104–14.
Vogeley, K., P. Bussfeld, A. Newen, S. Herrmann, F. Happé, P. Falkai, W. Maier. N. J. Shah, G. R. Fink, and K. Zilles. 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. NeuroImage 14.170–81.
Wellman, H. (1990). The child's theory of mind. Cambridge, MA: MIT Press.
Wimmer, H. and J. Perner. 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13.41–68.
Xu, J., S. Kemeny, G. Park, C. Frattali and A. Braun. 2005. Language in context: emergent features of word, sentence, and narrative comprehension. NeuroImage 25.1002–15.
Zysset, S., O. Huber, E. Ferstl and D. Y. von Cramon. 2002. The anterior frontomedian cortex and evaluative judgment: an fMRI study. NeuroImage 15.983–91.

