Neuropsychologia xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Neuropsychologia

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

Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception

Deborah Zaitchik[a,e,*], Caren Walker[f], Saul Miller[g], Pete LaViolette[h], Eric Feczko[i], Bradford C. Dickerson[b,c,d,e]

[a] Department of Psychiatry, Massachusetts General Hospital, USA
[b] Department of Neurology, Massachusetts General Hospital, USA
[c] Massachusetts Alzheimer's Disease Research Center and Frontotemporal Dementia Unit, Massachusetts General Hospital, USA
[d] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, USA
[e] Harvard Medical School, Boston, MA, USA
[f] University of California, Berkeley, USA
[g] Florida State University, Tallahassee, FL, USA
[h] Medical College of Wisconsin, Milwaukee, WI, USA
[i] Washington University, St. Louis, MO, USA

ARTICLE INFO

Article history:
Received 21 August 2009
Received in revised form 9 March 2010
Accepted 23 April 2010
Available online xxx

Keywords:
Functional magnetic resonance imaging
Theory of mind
Parietal cortex
Temporal cortex

ABSTRACT

By age 2, children attribute referential mental states such as perceptions and emotions to themselves and others, yet it is not until age 4 that they attribute representational mental states such as beliefs. This raises an interesting question: is attribution of beliefs different from attribution of perceptions and emotions in terms of its neural substrate? To address this question with a high degree of anatomic specificity, we partitioned the TPJ, a broad area often found to be recruited in theory of mind tasks, into 2 neuroanatomically specific regions of interest: Superior Temporal Sulcus (STS) and Inferior Parietal Lobule (IPL). To maximize behavioral specificity, we designed a tightly controlled verbal task comprised of sets of single sentences – sentences identical except for the type of mental state specified in the verb (belief, emotion, perception, syntax control). Results indicated that attribution of beliefs more strongly recruited both regions of interest than did emotions or perceptions. This is especially surprising with respect to STS, since it is widely reported in the literature to mediate the detection of referential states – among them emotions and perceptions – rather than the inference of beliefs. An explanation is offered that focuses on the differences between verbal stimuli and visual stimuli, and between a process of sentence comprehension and a process of visual detection.

1. Introduction

The development during childhood of a theory of mind – the ability to attribute mental states to others – has been characterized as involving both an early *referential* theory and a later *representational* one (Flavell, 1999; Perner, 1993; Wellman, Cross, & Watson, 2001) The early referential theory supports the ability to attribute desires, goals, perceptions, and emotions, an ability evident even in toddlers, while the later representational theory underlies the ability to attribute true and false beliefs, an ability not acquired until the preschool years.

1.1. The early referential theory of mind: desires, perceptions, emotions

By the time children are 2 years old, they spontaneously refer to *desires* (Bartsch & Wellman, 1995) and indicate their understanding that another person's desires may differ from their own (Repacholi & Gopnik, 1997). They speak about *perceptions* and *emotions* as well, referring to what people *see* and what people *feel*. Moreover, they speak about the causes of their emotions (Bretherton, Fritz, Zahn-Waxler, & Ridgeway, 1986; Wellman, Harris, Banerjee, & Sinclair, 1995), especially noting that frustrated *desires* lead to *anger* or *sadness* (Dunn & Brown, 2001; Lagattuta & Wellman, 2001). It has been claimed that such a coherent causal framework, one that allows prediction and explanation of human action, constitutes a genuine theory of mind (Wellman & Woolley, 1990). Still, it has been argued, this early theory does not require any understanding of mental *representation*. A goal or perception or emotion may be understood by the very young child as a *connection* of a particular sort toward an object or event (Flavell, 1988). For example, a *desire* for an object

* Corresponding author at: MGH Gerontology Research Unit, 149 13th St., Suite 2691, Charlestown, MA 02129, USA. Tel.: +617 726 5571; fax: +617 726 5760.
E-mail address: dzaitchiksamet@partners.org (D. Zaitchik).

0028-3932/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2010.04.031

may be understood as an *attraction*, a *pull*, that *refers* to the object. It would not be understood by the child as a *representation*. Indeed, a representational understanding of the mind is not evident until about age 4, when the child first succeeds in attributing a false belief to a deceived actor. The knowledge that beliefs can be *false* (that is, they can *misrepresent*) necessarily indicates the understanding that beliefs are *representations* (Wimmer & Perner, 1983; Wellman & Cross, 2001; Zaitchik, 1990). Different models of theory of mind have agreed that the ability to attribute a false belief represents a developmental watershed, and a good deal of current research is dedicated to specifying the precise cognitive mechanisms underlying this development (Leslie, 2000; Baron-Cohen, 1995). If new cognitive mechanisms come on-line to support the later theory of mind, an interesting question arises: is the neural circuitry subserving the attribution of the early referential mental states (*desires*, *perceptions*, *emotions*) different from the neural circuitry subserving the attribution of *beliefs*? Is there a neural substrate *specialized* for belief attribution? Rebecca Saxe and her colleagues have laid out two criteria that they consider necessary to support a claim that a particular brain region is *specialized* for the attribution of belief (Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004). First, the area must show an increased response to stimuli inviting *true belief* attributions as well as *false belief* attributions. Second, it must respond specifically to *belief* attributions. The present study focuses on the issue of specificity, directly comparing brain activity during attributions of beliefs vs. attributions of emotions and perceptions.

1.2. Neuroimaging studies of mental states: the role of the temporoparietal junction (TPJ)

Areas within the broad anatomic area termed TPJ have consistently been found to be active during tasks involving the attribution of mental states. A recent extensive meta-analysis (Van Overwalle, 2009) reports that 13 of the 15 studies involving attribution of *beliefs* showed activation of left TPJ (Ferstl & von Cramon, 2002; Wang, Lee, Sigman, & Dapretto, 2006), right TPJ (Grézes, Berthoz, & Passingham, 2006; Grèzes, Frith, & Passingham, 2004; Sommer et al., 2007; Wang et al., 2006), or bilateral TPJ (Gallagher et al., 2000; German, Niehaus, Roarty, Giesbrecht, & Miller, 2004; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe, Schulz, & Jiang, 2006). Importantly, these same regions have been shown to be *specific* to mental states: for example, they do not respond merely to the presence of a person or to non-mental representations such as an outdated photograph (Saxe & Kanwisher, 2003) and they do not respond to subjective sensations such as hunger or cold (Saxe & Powell, 2006). It seems likely then that any neural area responding preferentially to belief attribution will be located in the TPJ. Still, the TPJ is recruited during the attribution of *other* mental states as well – and these include *referential* states such as desires, intentions, and goals (Van Overwalle, 2009). Indeed, Van Overwalle (2009) has suggested that the TPJ is recruited for attribution of *all* temporary mental states. These would include perceptions and emotions as well as beliefs (it would exclude *enduring* properties, such as personality traits or long-standing dispositions). Thus, it is not clear whether there are regions within the TPJ that respond selectively to *representational* mental states (beliefs), as opposed to *referential* mental states (emotions and perceptions), or whether these regions are relatively unselective behaviorally.

This lack of specification is in part related to the design of tasks used in previous neuroimaging studies of theory of mind. To test the claim that different circuitry subserves early and late theory of mind, a neuroimaging study must directly compare the attribution of *referential* states (e.g., *emotions*, *perceptions*) with the attribution of *representational* states (e.g., *beliefs*). To date, however, there have been no published neuroimaging studies that *directly* compare beliefs with perceptions and emotions. Moreover, stimuli used in these paradigms vary widely, such that neuroimaging studies of *beliefs* have primarily used story stimuli, requiring the participant to *infer* the story character's belief from the causal events specified in the narrative (Ferstl & von Cramon, 2002; Gallagher et al., 2000; Perner et al., 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe, Moran, Scholz, Gabrieli, 2006; Saxe, Schulz, 2006; Vogeley et al., 2001). In contrast, studies of *emotions* and *perceptions* have generally used visual displays such as pictures of emotionally expressive faces (Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001) or videos of moving eyes (Pelphrey, Singerman, Allison, & McCarthy, 2003). These studies require visual *detection* of a mental state, a psychological process quite different from *inference*. Some studies mix the two. For example, one recent paper (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007) compares the areas recruited during two tasks – one involving stories about false beliefs, the other involving animated movements of geometric blobs. These tasks vary in many ways in addition to the specific types of mental states involved. Restricting the stimuli to stories would not completely solve the problem either: story tasks involve complex narratives whose length, causal structure, and variable content allow for differences among conditions that extend beyond the type of mental state. While previous studies have contributed in a variety of important ways to our understanding of the neural substrates of social cognition, it seems important to approach an investigation of perceptions, emotions, and beliefs using a tightly controlled behavioral paradigm. Such a study would use a within-subject design, a single type of stimulus (visual or verbal), a single type of psychological process (for example, target detection, causal inference, or sentence comprehension), and identical stimuli (save for the mental state to be attributed). This is the goal of the present study.

In addition to the issue of *behavioral specificity*, interpretation of the prior literature presents a challenge related to *anatomic specificity* as well. The TPJ is relatively large, encompassing inferior parietal lobule and superior temporal brain regions, and exhibits substantial anatomic variability between individuals. While previous analytic models often treated these regions as one, contemporary models suggest that they may have different functions. Specifically, the superior temporal cortex may be specialized for reading mental states from body cues; for example, regions of STS show increased activation in response to visual displays of eye, mouth, and hand movements (Allison, Puce, & McCarthy, 2000; David et al., 2008; Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005), as well as whole body movements (Saxe, Xiao, et al., 2004). The inferior parietal lobule (IPL), in contrast, may be responsible for mediating one's sense of *agency* – the feeling that the self is *causing* or *generating* an experience, not just passively undergoing it (Decety & Chaminade, 2003).

The issues outlined above have led some investigators to call for the use of developmentally motivated hypotheses, carefully defined regions of interest, and verbal stimuli in the design of future theory of mind neuroimaging studies (Saxe, 2006a, 2006b). The present study adopts this approach. As noted above, the developmentally motivated hypothesis is that one or more regions within the TPJ will respond selectively to *representational* mental states as opposed to *referential* states. With respect to anatomic specificity, our analytic approach employs the *a priori* use of anatomically defined regions of interest including the inferior parietal lobule and caudal superior temporal sulcus, defined at the individual subject level. The task uses verbal stimuli in the form of single sentences. Finally, to address the issue of behavioral specificity of the paradigm, the task conditions vary on only a single parameter: the type of mental state specified in the stimulus statement.

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

Table 1
Sample items from 4 conditions: Control (C), perception (P), emotion (E), belief (B).

Block	Condition		Statement	Question
1	C	1	It is likely that the nuts are rancid.	Is it likely that the nuts are fresh?
1	C	2	It is now that the tide is coming in.	Is it only later that the tide is coming in?
1	C	3	…	…
1	C	4	…	…
1	C	5	…	…
1	C	6	…	…
2	P	1	She hears that the tide is coming in.	Does she hear that the tide is going out?
2	P	2	He tastes that the nuts are rancid.	Does he taste that the nuts are rancid?
2	P	3	…	…
2	P	4	…	…
2	P	5	…	…
2	P	6	…	…
3	E	1	He is furious that the nuts are rancid.	Is he furious that the nuts are fresh?
3	E	2	He is afraid that that the tide is coming in.	Is he happy that the tide is coming in?
3	E	3	…	…
3	E	4	…	…
3	E	5	…	…
3	E	6	…	…
4	B	1	He remembers that the nuts are rancid.	Does he remember that the nuts are fresh?
4	B	2	He thinks that the tide is coming in.	Does he very much doubt that the tide is coming in?
4	B	3	…	…
4	B	4	…	…
4	B	5	…	…
4	B	6	…	…

2. Materials and methods

2.1. Participants

Participants were 15 right-handed, native English speakers (8 males, 7 females; mean age = 22.4 years, range of 20–24; mean education = 15.7 years, range of 11–18 years) who were recruited via local advertisements and received $100 each for participation. A standard health screen was administered over the phone to ensure no history of medical, psychiatric, or neurological illness prior to enrollment. Upon arrival to the testing center, all participants were administered a handedness screen (the Edinburgh Handedness Survey), and a reading test (the American National Adult Reading Test) to further ensure eligibility. All participants were native English speakers who gave written informed consent prior to their inclusion in the study. The study was approved by the Partners Healthcare System Human Research Committee and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

2.2. Stimuli and cognitive task

2.2.1. The paradigm: propositional attitudes

Using a verbal paradigm, our study provides a direct comparison of reasoning about *beliefs*, a representational state, and reasoning about *emotions* and *perceptions*, two referential states, in the same participants. In order to directly compare these different mental states in a single paradigm, the present study takes advantage of the fact that beliefs, perceptions, and emotions can all serve as *propositional attitudes*. Consider the following sentences: (1) She *thinks* that the door is open; (2) She *sees* that the door is open; (3) She *is afraid* that the door is open. Each sentence has the same Agent ('She') and the same embedded Proposition ('that the door is open'). The only difference in these sentences is the Attitude that the Agent takes to the proposition – in one case a *belief*, in one case a *perception*, and in one case an *emotion*. In looking at the neural activations underlying comprehension of these sentences, then, we can directly compare the processes involved in comprehension of a belief – a representational state – to those involved in comprehension of a perception or an emotion – two referential states – using *precisely* the same stimuli. Preferential activation of any brain region during the *belief* condition – but neither of the other two conditions – would provide strong evidence that this brain region makes a specific contribution to the representational theory of mind.

The sentence stimuli were categorized according to four conditions – 3 conditions involving different types of mental state (BELIEF, EMOTION, and PERCEPTION conditions) and 1 syntax CONTROL condition. The CONTROL condition was included to enable us to identify neural areas that might be activated by sentential complement syntax alone, even in the absence of a mental state attribution. Sentences in each condition were identical to those in every other condition except for the predicate which indicated type of mental state (or, in the CONTROL condition, the absence of a mental state). For example, in the BELIEF condition, the subject would be presented with 'He *believes* that the cat is on the couch'; in the EMOTION condition, 'He *is furious* that the cat is on the couch'; in the PERCEPTION condition, 'He *sees* that the cat is on the couch.; in the syntax CONTROL condition, '*It is true* that the cat is on the couch'. Specific terms used in each condition were as follows: (a) BELIEF: *remembers*, *believes*, *thinks*, *knows*; (b) PERCEPTION: *hears*, *sees*, *tastes*, *smells*; (c) EMOTION: *furious*, *happy*, *sad*, *afraid*; (d) CONTROL: *is now*, *is likely*, *is true* (used twice). Sentential complements were carefully chosen so as to avoid blurring the distinction between representational and referential verbs. For example, instead of the sentence 'She saw *that he had a point*' (where the meaning of *saw* may be identical to *understood* or *thought*), participants were presented with 'She saw that *the room was yellow*' (where *saw* functions more clearly as a perceptual term).

To ensure that participants attended to and fully comprehended the entire sentence – both mental state verb and sentential complement – each sentence was followed by a simple corresponding yes/no question. *Yes* questions generally repeated the information given in the statement, but using a synonymous verb (for example, after hearing the sentence, 'He is *furious* that the cat is on the couch', the subject is asked, 'Is he *angry* that the cat is on the couch?'); *No* questions either denied the assertion of the statement, using an antonym or a negation (e.g., 'Is he *happy* that the cat is on the couch?') or negated the statement by changing something in the sentential complement phrase ('Is he furious that the cat is *on the floor*?').

Stimuli were administered in a block design (see Table 1 for sample items) by condition. Each block consisted in 6 statement/question pairs. Each statement was presented alone for an initial 3 s, and was followed by a corresponding question (presented below the statement) for an additional 4 s. Blocks were separated by a fixation cross of 24 s. There were 8 blocks per run (2 of each condition), and 4 runs in total. In sum, then, each participant was presented with 192 statement/question pairs (6 statement/question pairs per block × 8 blocks per run × 4 runs). Answers were coded by button press for "yes" or "no" to each question based on the statement provided. All subjects were trained using a comparable practice run outside of the scanner immediately prior to testing.

The order of the presentation for each block type was counterbalanced across the four runs, as well as within each individual run (with the last four blocks in each run presented in opposite order from the first four). All of the material presented in the paradigm was assessed prior to selection via a pilot study of twenty healthy subjects of comparable age and education during individual offline testing sessions.

All stimuli were presented on a PC laptop using Eprime software (Psychology Software Tools, Inc, Pittsburgh, PA). Stimuli were projected into the scanner using a rear mounted LCD projector in conjunction with a mirror mounted on the head coil.

2.3. Functional imaging

A Siemens (Siemens Medical Solutions, Malvern, PA) 1.5T Magnetom Avanto system equipped with a Total Imaging Matrix (TIM) 12-channel head coil was used to acquire high-resolution T1-weighted anatomical data (MPRAGE: TR/TI/TE 2730/1000/3.31 ms), FOV = 256, FA = 7°, 128 sagittal slices, thickness = 1.33 mm,

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

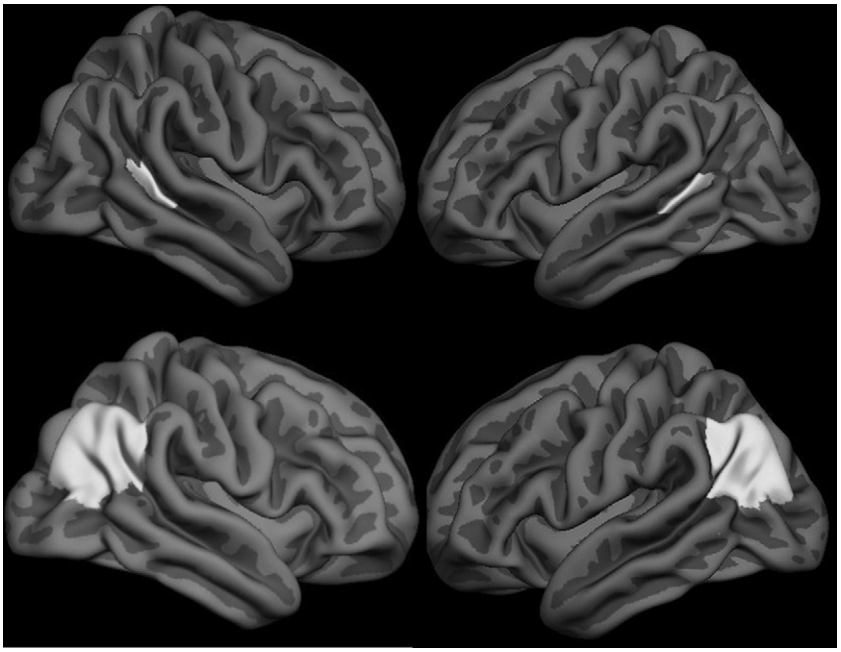

Fig. 1. The two regions of interest employed in this study representing the temporoparietal junction (TPJ), (A) The caudal superior temporal sulcus (STS) and (B) the inferior parietal lobule (IPL). For illustration purposes, these regions of interest are shown on the left and right lateral surfaces of an average representation of the cortical surface from forty subjects, but they were defined for the analyses in this study at the individual subject level.

matrix 192 × 256 (1.3 mm × 1 mm in-plane resolution), and T2*-weighted gradient-echo echo-planar (EPI) functional data (TR = 2000 ms, TE = 30 ms, 30 axial slices, 4-mm thickness, 0.4-mm interslice skip, 192 mm FOV, matrix 64 × 64 (3 mm × 3 mm in-plane resolution)). Four additional volumes were collected and discarded at the beginning of each run to allow for T1 equilibration. An on-line automatic slice-positioning atlas was used for slice prescription (Van der Kouwe et al., 2005).

2.4. Behavioral data analysis

Accuracy and time to respond to the question that followed each stimulus statement were recorded. The effect of mental state condition on each of these measures was analyzed by ANOVA.

2.5. Functional-anatomic ROI-based fMRI data analysis

Data were preprocessed using Statistical Parametric Mapping (SPM2; Wellcome Department of Cognitive Neurology, London, UK) for Matlab (The Mathworks, Inc, Natick, MA). Functional data were realigned using INRIAlign, a motion-correction algorithm unbiased by local signal changes. The functional data were then smoothed using a 8 mm Gaussian kernel, and modeled with the canonical hemodynamic response function. A high-pass filter of 240 s was used to remove low-frequency signal (e.g., drifts across run). Data remained in native space of 3 mm × 3 mm × 3.6 mm resolution and were then co-registered to individual high-resolution structural images.

A functional-anatomic analysis procedure was used to test a priori hypotheses regarding the localization of functional activation within two temporoparietal regions of interest in each hemisphere. Regions of interest (ROI) were derived from each individual subject's high-resolution T1 structural scan using a semi-automated anatomic reconstruction and labeling procedure using Freesurfer software (http://surfer.nmr.mgh.harvard.edu/) (Desikan et al., 2006). Boundaries implemented on the cortical surface model were verified by a manual operator as correctly reflecting the manual anatomic protocol used to define them. The first ROI (inferior parietal lobule, or IPL) was an inferior parietal cortex region that includes the angular gyrus and lies inferior to the superior parietal lobule. The rostral boundary, not included in this ROI, was the supramarginal gyrus and the caudal boundary was the lateral occipital cortex. The medial and lateral boundaries were the superior parietal lobule and the middle temporal gyrus, respectively. The second ROI (caudal superior temporal sulcus, or STS) was the caudal third of the superior temporal sulcus. The caudal boundary was the last slice where the superior temporal gyrus separates from the rest of the cortex (posterior to becoming continuous with the supramarginal gyrus). The dorsal and ventral boundaries were the crowns of the superior and middle temporal gyri, respectively. See Fig. 1 for example regions of interest.

Each subject's motion-corrected echo-planar data were co-registered to that subject's T1 data. Cortical surface-based ROI labels were then resampled into each subject's native volume space. Percent signal change and beta weight values were extracted for all conditions from all voxels within the ROIs meeting an uncorrected functional threshold of $p < 0.05$ based on the [ANY MENTAL STATE CONDITION] > CONTROL contrast. This approach selects for voxels that are involved in mental state inference but is unbiased with respect to the questions of interest in this study, namely differential activation for beliefs, emotions, or perceptions. Data were extracted using the SPM MarsBaR toolbox (http://marsbar.sourceforge.net/). Statistical analyses were performed (SPSS 13.0, SPSS Inc., Chicago, IL) using analysis of variance to compare condition-specific activity measures within ROIs. Conditions included BELIEF, PERCEPTION, EMOTION, CONTROL, and FIXATION. For condition-specific ROI analyses of fMRI data, one-way ANOVAs were used with post hoc comparisons.

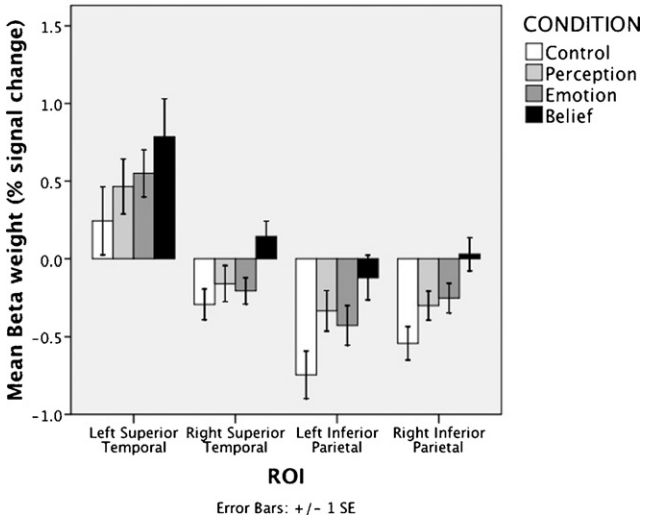

Fig. 2. Beta weights from each task condition from within each region of interest, including left caudal superior temporal sulcus, right caudal superior temporal sulcus, left inferior parietal lobule, and right inferior parietal lobule. Within each region of interest, the color bars represent the beta weight of the parameter estimate of the BOLD signal during each condition (vs. fixation), including CONTROL (CONT), EMOTION (EMOT), BELIEF (BEL), and PERCEPTION (PERC). Error bars represent one standard error of the mean.

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

Table 2
Results from whole brain analyses (*x*, *y*, *z* represent MNI coordinates).

x	*y*	*z*	Cluster size	*Z* value	Region
Belief vs. control					
26	−76	2	189	3.61	Right Lingual Gyms
−6	32	4	70	3.18	Left Anterior Cingulate
−10	−56	22	180	3.06	Left Posterior Cingulate
−52	−74	26	114	2.89	Left Superior Temporal Gyms
−20	0	30	21	2.74	Left Caudate Nucleus
48	−58	18	108	2.65	Right Superior Temporal Gyms
−8	56	−4	30	2.64	Left Superior Frontal Gyms
54	−34	24	61	2.61	Right Inferior Parietal Lobule
Emotion vs. control					
−8	46	54	56	3.15	Left Superior Frontal Gyms
26	−46	8	57	2.82	Right Parahippocampal Gyms
−10	58	40	36	2.67	Left Superior Frontal Gyms
−66	−20	−6	19	2.62	Left Middle Temporal Gyms
−16	−84	34	18	2.6	Left Cuneus
Perception vs. control					
−36	−42	8	195	3.83	Left Caudate Nucleus
38	−82	2	331	3.64	Right Middle Occipital Gyms
54	−66	2	104	3.24	Right Inferior Temporal Gyms
−52	−76	26	71	2.95	Left Middle Temporal Gyms
−2	34	−2	63	2.88	Left Anterior Cingulate Gyms
14	66	14	94	2.61	Right Superior Frontal Gyms
Emotion vs. belief					
62	−32	18	50	3.6	Right Superior Temporal Gyms
−64	−24	24	125	3.45	Left Inferior Parietal Lobule
56	4	8	26	2.99	Right Precentral Gyms
Belief vs. emotion					
−10	56	34	89	3.46	Left Superior Frontal Gyms
−34	−78	20	74	3.11	Left Middle Temporal Gyms
−28	−76	38	84	2.77	Left Precuneus
Belief vs. perception					
−44	26	4	142	3.5	Left Inferior Frontal Gyms
20	44	42	128	2.81	Right Superior Frontal Gyms
Perception vs. belief					
−22	44	−6	38	3.12	Left Middle Frontal Gyms
−52	−64	0	46	2.77	Left Inferior Temporal Gyms

3. Results

3.1. Behavioral data[1]

One-way ANOVA revealed significant effects of condition on both accuracy [*F*, df(3, 2863) = .003] and reaction time [*F*, df(3, 2863) = .001]. Post hoc tests show that these effects were due solely to the syntax CONTROL condition, which led to less accurate responses than the EMOTION and PERCEPTION conditions (*p*s < 0.02) and slower responses than the BELIEF and EMOTION conditions (*p*s < 0.01). There were no significant differences in either accuracy or reaction time among the three mental state conditions.

3.2. Functional imaging

Functional-anatomic ROI analyses for a priori regions were performed initially at the individual subject level to optimize localization of activation data with respect to anatomic landmarks. Visualization of individual subject statistical maps overlaid on both mean individual subject EPI and T1 data revealed good correspondence of landmarks, indicating that the quantitative metrics accurately represent BOLD signal effects within the ROIs.

For each ROI, the estimated BOLD signal in 5 conditions (BELIEF, PERCEPTION, EMOTION, CONTROL, FIXATION) was analyzed by one-way ANOVA plus post hoc comparisons (Fig. 2). All differences reported here are significant at $p < 0.05$ except for those reported as trends, which are significant at $p < 0.1$. For the right IPL [$F(4) = 6.8$], greater activation was present in the BELIEF condition than in the EMOTION, PERCEPTION, and CONTROL conditions. EMOTION AND PERCEPTION were also different from CONTROL (they were less deactivated) but they were not different from each other. EMOTION, PERCEPTION, and CONTROL conditions were different from FIXATION, but the BELIEF condition was not. The left IPL [$F(4) = 5.4$] showed greater activation to BELIEF than to CONTROL, and a trend toward greater activation to BELIEF than to EMOTION ($p < 0.1$). CONTROL, EMOTION, and PERCEPTION were all significantly deactivated compared to FIXATION, while BELIEF was not.

Right STS [$F(4) = 3.8$] showed greater activation in the BELIEF condition compared to EMOTION, PERCEPTION, and CONTROL, which were not different from one another. Only CONTROL was different from FIXATION. For the left STS [$F(4) = 2.8$, BELIEF was different from CONTROL, EMOTION, and PERCEPTION which all showed greater activation than FIXATION.

In addition to the analysis of the primary a priori regions of interest described above, additional a priori regions of interest were investigated based on previous studies of theory of mind. These regions of interest were the Precuneus and the Medial Orbital Frontal ROIs defined as described previously (Desikan et al., 2006). The ROI results from these regions indicated the presence of the predicted deactivation below baseline during task conditions but the absence of any statistically significant effects (including trend-level effects) when the conditions were compared to each other.

[1] Due to a computer glitch, some behavioral data were lost. These were replaced by behavioral data of several pilot subjects.

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

In addition, a whole brain analysis was run in SPM to further explore the results, using a p value of <0.005 uncorrected, and an extent threshold of 10 voxels. The results of this exploratory analysis are presented in Table 2.

The exploratory analysis highlights a few unanticipated findings that deserve further investigation in future studies. For example, there are two clusters of greater activation for EMOTION than BELIEF near the temporoparietal ROIs of primary interest in this study–Right STG and Left IPL. However, careful inspection of the peaks of these clusters indicates that they are in zones outside of the a priori ROIs. That is, the area in the right superior temporal gyrus that is more active in EMOTION than BELIEF is in the dorsal convexity of the posterior superior temporal gyrus, not in the STS. The region of the IPL that shows similar modulation is in the supramarginal gyrus rostral to the a priori ROI.

4. Discussion

This study was motivated by a very puzzling developmental phenomenon – a 2-year lag between the toddler's ability to attribute *referential* states (e.g., *emotions and perceptions*) and the preschool child's ability to attribute *representational* mental states (e.g., *beliefs).* As noted above, Saxe and Wexler (2005) posed the question of whether representational mental states differentially activate selective regions of the neural substrate. If so, perhaps developmental constraints on brain function would provide an explanation for lags in conceptual development in theory of mind.

The recent interest in neuroimaging studies of social cognition has led to the identification of a number of brain regions which are generally thought to be implicated in theory of mind. As mentioned above, and as is clear from van Overwalle's recent review (2009), the TPJ is quite consistently implicated in inference of *representational* mental states such as *beliefs* (Ferstl & von Cramon, 2002; Gallagher et al., 2000; German et al., 2004; Grezes, Frith, & Passingham, 2004; Grezes, Berthoz, & Passingham, 2006; Perner et al., 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe, Moran, 2006; Saxe, Schulz, 2006; Sommer et al., 2007; Wang et al., 2006). For example, Saxe and Powell (2006) found bilateral activation of TPJ to stories about *beliefs* but not to stories about *nonrepresentational* internal states (e.g., being tired or hungry). From these and other findings (Saxe, Carey, & Kanwisher, 2004; Saxe, Xiao, et al., 2004), they argue that the early- and late-developing components of theory of mind rely on separate psychological and neural mechanisms, that the TPJ is the substrate of the late-developing *representational* theory of mind, and that these mechanisms remain distinct into adulthood. In line with this, Gobbini et al. (2007) report that false belief stories, but not animations of geometric shapes depicting social interactions, activated a neural circuit that included the TPJ. Like Saxe and colleagues, they argue that the representation of *beliefs* and the representation of *intentions*, another type of referential mental state represented in the early-developing theory of mind, involve distinct neural systems. Though these findings do not investigate the neural circuitry underlying attributions of *perceptions* and *emotions* – the specific referential mental states we sought to investigate in our study – they are certainly in line with a claim of separate underlying networks for representational and referential states.

In contrast to beliefs, the attribution of referential states such as *perceptions* (Pelphrey et al., 2003) and *emotions* (Narumoto et al., 2001), *intentions* and *goals* (Allison et al., 2000) have all been shown to recruit pSTS. For this reason, pSTS has been characterized as the substrate for understanding *goal-driven* action (Saxe, Carey, et al., 2004; Saxe, Xiao, et al., 2004). For example, social interactions implied simply by the movement of geometric shapes (Gobbini et al., 2007) activated an extensive region that included pSTS, the frontal operculum, and inferior parietal lobule (IPL) – a system that the authors consider to be the substrate underlying *action* understanding. There is a good deal of evidence that pSTS mediates the understanding of *intentional action*. This will be discussed further below. It is worth noting here, however, that these are precisely the types of mental states attributed by toddlers.

4.1. A new paradigm

One might thus be tempted to infer that attribution of *representational* mental states is subserved by the TPJ while attribution of *referential* mental states is subserved by the pSTS. This seems premature. The problem is that the evidence is indirect, collapsing across studies with very different tasks, very different types of stimuli, and very different cognitive processes. These range from recognizing happiness in a smiling face to noticing a change of gaze in a pair of moving eyes to reading a goal or intention in someone's approach toward an object, to inferring someone's false belief from a story narrative. These are, indeed, all theory of mind tasks; they are all part and parcel of social cognition. Still, the differences in paradigms contribute many sources of noise that caution against using them as direct comparisons of the neural circuitry underlying attribution of different mental states.

4.2. Verbal stimuli

Saxe's (2006a) suggestion to use verbal stimuli is certainly well taken; story studies reduce the problems caused by contrasting different processes. Still, as remarked above, stories themselves are not devoid of noise. They vary in length, content, causal connections, etc. Taking the argument in favor of using verbal stimuli one step further, we designed a task that varied conditions of interest *only* by the type of mental state attributed. Among the strengths of this method is the increased confidence that a difference in the neural activation is due to a difference in type of mental state. Second, the noise reduction makes the method more sensitive to differences that might have been indistinct in a noisier paradigm. We recognize, however, that such a narrow paradigm has its limitations. Our task is confined to simple sentence comprehension; it does not demand detection from visual cues, a frequent source of mental state information in the real world, nor does it demand inferences from relevant information. Indeed, our stimuli explicitly *provide* the mental state; the subject need only *comprehend* the sentence (a requirement that our simple follow-up questions assure us has been met). We reasoned that the greater generalizability of a more varied presentation was worth sacrificing in the interest of the high specificity and sensitivity of our task design. This strategy has led us to the following findings.

1. First, with respect to STS, we found preferential activation to *beliefs* as compared to *emotions and perceptions* (marked on the right and with a trend in the same direction on the left). Given the previous literature, this is surprising. As noted above, previous studies indicate that the STS (especially right STS), is particularly activated during the attribution of *referential* states – in particular *perceptions* and *emotions* (Narumoto et al., 2001; Pelphrey et al., 2003). Indeed, the TPJ – the region generally associated with *belief* attribution – is often specifically *contrasted* with STS (Saxe, 2004b). In the present study, however, not only did the BELIEF condition more strongly activate bilateral STS than did PERCEPTION and EMOTION, but right STS actually showed *deactivation* (compared to fixation) to both PERCEPTION and EMOTION conditions, precisely the conditions that should have led to increased activation. What could explain this?

A closer look at the literature cited earlier suggests the explanation. As noted above in arguing for a tightly controlled paradigm, activation of STS (especially right STS) during attribution of ref-

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

erential states has generally been found in the context of *visual* displays of parts of the body. *Perceptions*, for example, are detected in visual displays of people's eye movements (Pelphrey et al., 2003, 2005), goals are detected in visual displays of hand movements or whole body movements (Allison et al., 2000; Pelphrey et al., 2005; Saxe, Xiao, et al., 2004), emotions are detected in visual displays of facial expressions (Narumoto et al., 2001). Indeed, Saxe, Xiao, et al., 2004 characterize this region as the substrate of *observed* goal-driven action. If the mechanisms involved in this processing are specifically geared to *visual* input, then – in the absence of visually presented body movements, in the presence of *verbal* attributions alone – emotions and perceptions may only weakly activate (left STS) or even deactivate (right STS) this region. This is understandable in light of the non-human primate anatomic connectivity data indicating that the STS is a cortical region responsible for high-level *visual* integration (Pandya & Yeterian, 1985). *Beliefs*, on the other hand, fall in a different category; they cannot be visually detected and interpreted on the basis of bodily action. Though beliefs may, in principle, be inferred from evidence about a person's exposure to information, a *visual* display of that exposure is unlikely to be available, unlikely to be complete, and unlikely to be efficient. On the contrary, *verbal* information should more quickly and reliably communicate information about beliefs and *verbal* stimuli should prove stronger activators of the neural substrate underlying belief ascriptions. These ideas are speculative, however, and deserve systematic investigation.

2. Another interesting result of this study involves the IPL. Though bilateral IPL was more activated by beliefs than by perceptions and emotions, it was equally activated by fixation. This finding raises the possibility that there is something about belief attribution that is similar to the resting state, when the brain is not engaged in any directed task. According to Buckner and Carroll (2006), this undirected mode of processing, sometimes referred to as the *default mode* (Raichle et al., 2001), involves a highly stereotypic pattern of activity that characterizes a state when people passively think about themselves and are relatively inattentive to their immediate surroundings. The default network, they suggest, mediates several cognitive functions involving self-projection – thinking about the past, thinking about the future, and thinking about other people's beliefs. On this view, the underlying similarity of these functions is the demand to detach from one's own immediate sensorimotor reality and to consider different perspectives. This, however, raises the question of why emotions and perceptions, which also vary with different perspectives, do not show the same pattern of activation. We think it possible that beliefs are taken as *models of reality* in a different sense from emotions and perceptions. Paradigmatically, perhaps, they come ready for decoupling from reality, wearing their representational nature on their sleeve in a way that emotions – and even perceptions – do not. Indeed, it could be that the neurocognitive processing for beliefs engages, by default, a decoupling process that is not an element of emotion and perception processing. Such an argument, if possible to operationalize, would go hand in hand with the developmental finding of the 2-year lag between attribution of emotions and perceptions and the attribution of beliefs. Perceptions (in particular) *can* function as mental representations, they *can* misrepresent reality much like beliefs, and it is worth noting that the child's ability to attribute conflicting perceptions is acquired at about the same time as the ability to attribute conflicting beliefs (Flavell, 1999). Indeed, it is for that reason that one might have expected *perception* to pattern with *belief* in adults such as our study participants, even if they function as referential states in toddlers, as developmental psychologists have claimed. Interestingly, however, perceptions do not pattern with beliefs; they do not modulate the neural substrates responsible for belief attribution in the same manner as beliefs themselves. Of course, our task did not require the processing of *conflicting* emotions, or *conflicting* perceptions, but neither did it require the processing of *conflicting* beliefs. The suggestion that belief processing is similar to default mode processing is quite speculative; a better explanation for the similar activation profiles in bilateral IPL may lie in the task demands. Still, this is an issue which deserves further investigation.

3. There is another sense in which the findings of the present study offer *particularly* strong support for the claim (Saxe, Carey, et al., 2004; Saxe, Xiao, et al., 2004) that the neural substrate of belief attribution remains distinct even in adulthood. Specifically, the use of the syntactic complement structure, even when used in an emotion attribution, may *logically* entail a belief attribution. Consider, for example, the sentence 'He is happy that the cat is on the couch'. No one could be *happy* [that the cat is on the couch] without *believing* [that the cat is on the couch]. Similarly, the statement 'She *sees* that her sister is in the kitchen' is generally taken as first-rate evidence for 'She *thinks* that her sister is in the kitchen'. Nevertheless, despite the easy representational readings invited by our stimuli, emotions and perceptions did not activate our ROI's as beliefs did.

4. With respect to the regions of interest of primary focus in the present study, the exploratory whole brain analysis was generally supportive of the ROI analyses. It highlighted 2 areas that we did not anticipate which deserve further investigation in future studies. Specifically, there are 2 clusters of greater activation for EMOTION than BELIEF that are near our ROI's – Right STG and Left IPL. However, careful inspection of the peaks of these clusters indicates that they are in zones outside of our ROIs; that is, the area in the right superior temporal gyrus that is more active in EMOTION than BELIEF is in the dorsal convexity of the posterior superior temporal gyrus, not in the STS. The region of the IPL that shows similar modulation is in the supramarginal gyrus rostral to our ROI. Further investigation is warranted of the possibility that nearby regions within the temporoparietal junction are modulated differentially as part of emotional processing vs. mental state inference. One reasonable approach to take to further investigate these effects would be similar to a prior study comparing activation during a false belief paradigm vs. a self-reflection paradigm, which suggested substantial overlap between the medial prefrontal and medial parietal/precuneus regions for the two paradigms but a relatively greater extent of the TPJ that was recruited during false belief than self-reflection (Saxe, Moran, 2006).

5. The present task involves the attribution of beliefs that are either marked as true (e.g., 'He *knows* that the room is painted yellow') or beliefs that are unmarked with respect to truth (e.g., 'He *thinks* that the room is painted yellow'). This is different from many studies which, following the developmental literature, have specifically focused on the attribution of *false* beliefs. As noted above, understanding false beliefs demands understanding beliefs as *representations* – necessarily, since they must be understood as *mis*representations. This is not true of true beliefs. It therefore seems possible, at least, that they do not require the same sort of processing as false beliefs, and that they do not recruit the same neural regions. Since our study does not specifically contrast true beliefs with false beliefs, it cannot speak directly to the differences in these conditions; however, it can indicate whether true belief attribution and unmarked belief attribution recruit either of our regions of interest. As it turns out, they recruit both – and significantly more than do the referential states with which they are compared. This is just what one would expect *if true or unmarked beliefs were treated as representational states*.

Claims in the literature about the neural substrate of true vs. false beliefs are relatively few and somewhat conflicting. Saxe and Kanwisher (2003) reported that identical voxels in individual subjects responded to both true and false beliefs, particularly strong evidence for a shared neural substrate even in adulthood. In line with this, Jenkins and Mitchell (2009) recently reported that the TPJ subserves inferences about beliefs – true and false – while

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

other regions (e.g., MPFC) may be more sensitive to *evaluating* beliefs with respect to truth. These results might be contrasted with those recently reported in a study directly comparing true and false beliefs in a story format. According to Sommer et al. (2007), the TPJ is recruited during attribution of false – but not true – beliefs. It is the IPL, they report, that is recruited for both true and false beliefs. These authors suggest that the TPJ is essentially a component for *decoupling misrepresentations* (such as false beliefs) from one's primary representation of reality, a critical function for a representational theory of mind. Similarly, Frith and Frith (2006) have characterized the TPJ as specifically involved in the attribution of multiple or conflicting representations (those that play a role in false belief or perspective-taking). It seems likely that some of the discrepancies found in the literature are due to differences in paradigms while others may be due to differences in the anatomic specificity of the regions of interest. After all, Sommer, Döhnel, Meinhardt, & Hajak (2008) report activation of IPL *in contrast* to activation of TPJ, while we consider IPL to be *part of* TPJ. Clearly, more direct comparisons involving true and false beliefs will be needed to clarify this issue.

The present study was designed to contribute a rigorous test of the hypothesis that the early referential theory of mind and the late representational theory of mind recruit distinct regions in the brain. Evidence that our 2 regions of interest within TPJ were recruited for sentences involving attributions of *belief* – and NOT for virtually identical sentences involving attributions of *emotion* or *perception* – contributes strong support for the claim that belief attribution has its own neural substrate. Establishing that this is the case is an important step in understanding the 2-year developmental lag between early and late theory of mind, raising the question of whether late theory of mind is late because it must await development of neural substrate that is not necessary for early theory of mind. For example, might belief attribution require greater connectivity between IPL/STS and the slower developing frontal regions than attribution of perception or emotion?

Of course, the tightly controlled paradigm which provided for the rigor of the investigation had its costs. For one thing, our use of propositional attitudes and sentential complement structures, while allowing us to employ virtually identical stimuli across conditions, did not allow us to include *desires* and *goals* among the referential states under investigation. Paradigms that could incorporate a fuller range of referential mental states – but still maintain tight control across conditions – would provide a more complete picture. We're confident that the intense interest in neuroimaging and theory of mind, an interest which has led to a large and thriving domain of research, will lead to such investigations soon.

Acknowledgements

This study was supported by the NIA (K23-AG22509, R01-AG29411), the NCRR (P41-RR14075, U24-RR021382), and the Mental Illness and Neuroscience Discovery (MIND) Institute. The authors thank Mary Foley, Larry White, and Jill Clark for technical assistance.

References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Science*, *4*, 267–278.
Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.
Bartsch, K., & Wellman, H. M. (1995). *Children Talk About the Mind*. London/New York: Oxford University Press.
Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotion: A functionalist perspective. *Child Development*, *57*, 529–548.
Buckner, R. L., & Carroll, D. C. (2006). Self-projection and the brain. *Trends in Cognitive Science*, *11*(2), 49–57.
David, N., Aumann, C., Santos, N. S., Bewernick, B. H., Eickhoff, S. B., Newen, A., et al. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Social Cognitive and Affective Neuroscience*, *3*, 279–289.
Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, *12*, 577–596.
Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980.
Dunn, J., & Brown, J. (2001). Emotion, pragmatics and developments in emotion understanding in the preschool year. In D. Bakhurst, & S. Shanker (Eds.), *Jerome Bruner: Language, Culture, Self* (pp. 88–103). Thousand Oaks, CA: Sage.
Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *NeuroImage*, *17*, 1599–1612.
Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 244–267). New York/London: Cambridge University Press.
Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review Psychology*, *50*, 21–45.
Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*, 531–534.
Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia*, *38*, 11–21.
German, T., Niehaus, J. L., Roarty, M. P., Giesbrecht, B., & Miller, M. B. (2004). Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *Journal of Cognitive Neuroscience*, *16*, 1805–1817.
Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind F tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803–1814.
Grézes, J., Berthoz, S., & Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *NeuroImage*, *30*, 601–608.
Grèzes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage*, *21*, 744–750.
Grezes, J., Frith, C., & Passingham RE. (2004). Brain mechanisms for inferring deceit in the actions of others. *Journal of Neuroscience*, *24*, 5500–5505.
Grezes, J., Berthoz, S., & Passingham RE. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *Neuroimage*, *30*, 601–608.
Jenkins, A. C., & Mitchell, J. P. (2009). Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex*, *21*(8), 1560–1570.
Lagattuta, K. H., & Wellman, H. M. (2001). Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion. *Child Development*, *72*, 82–102.
Leslie, A. (2000). "Theory of Mind" as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press, pp. 1235–1247
Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cognitive Brain Research*, *12*, 225–231.
Pandya, D. N., & Yeterian, E. H. (1985). Architecture and connections of cortical association areas. In A. Peters, & E. G. Jones (Eds.), *Cerebral Cortex* (pp. 3–62). New York: Plenum Press.
Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An FMRI study of eye, mouth and hand movements. *Cerebral Cortex*, *15*, 1866–1876.
Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia*, *41*, 156–170.
Perner, J. (1993). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The role of left and right temporo-parietal junction. *Social Neuroscience*, *1*, 245–258.
Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*, 676–682.
Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*, 12–21.
Saxe, R. (2006a). Why and how to study theory of mind with fMRI. *Brain Research*, *1079*, 57–65.
Saxe, R. (2006b). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*, 235–239.
Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroanatomy. *Annual Review of Psychology*, *55*, 87–124.
Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*, 1835–1842.
Saxe, R., Moran, J. M., Scholz, J., & Gabrieli. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Journal Social Cognitive and Affective Neuroscience*, *1*(3), 229–234.
Saxe, R., & Powell, L. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692–699.

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.04.031

Please cite this article in press as: Zaitchik, D., et al. Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia* (2010), doi: