Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts

Nir Jacoby & Evelina Fedorenko

Routledge
Taylor & Francis Group

Check for updates

REGULAR ARTICLE

Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts

Nir Jacoby [a] and Evelina Fedorenko [b,c,d]

[a]Department of Psychology, Columbia University, New York, NY, USA; [b]Harvard Medical School, Boston, MA, USA; [c]Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA, USA; [d]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

ABSTRACT

In addition to understanding individual word meanings and processing the syntactic and semantic dependencies among those words within a sentence, language comprehension often requires constructing a higher-order discourse structure based on the relationships among clauses and sentences in the extended context. Prior fMRI studies of discourse-level comprehension have reported greater activation for texts than unconnected sentences in what-appear-to-be regions of the Theory of Mind (ToM) network. However, those studies have generally used narratives rich in mental state content, thus confounding coherence and content. We report an fMRI experiment where ToM regions were defined functionally in each participant, and their responses were examined to texts vs. sentence lists. Critically, we used expository texts to minimise mental state content. Medial frontal but not posterior ToM regions exhibited small but reliable increases in their responses to texts relative to unconnected sentences, suggesting a role for these regions in discourse comprehension independent of content.

Introduction

Lake Baikal is a rift lake in the south of the Russian region of Siberia. It is the largest (by volume) freshwater lake in the world, containing roughly 20% of the world's unfrozen surface fresh water. Lake Baikal was formed as an ancient rift valley, having the typical long crescent shape. Baikal is home to more than 1,700 species of plants and animals, two-thirds of which can be found nowhere else in the world. (Edited from Wikipedia https://en.wikipedia.org/wiki/Lake_Baikal)

When reading the text above, you were able to integrate information from the different sentences into a coherent mental representation (e.g. Grosz & Sidner, 1986; Hobbs, 1985; Kintsch, 1998; Marcu, 2000; Zwaan, Langston, & Graesser, 1995). The first sentence introduced the topic (i.e. Lake Baikal), and subsequent sentences provided further information, thus enriching the evolving representation. Furthermore, having read the text, you are now equipped with knowledge to make new inferences about Lake Baikal even if you never heard about it before. What cognitive and neural processes enable the construction of these rich and complex representations during text comprehension?

Language comprehension has long been associated with a left-lateralised network of frontal, temporal, and parietal brain regions (e.g. Bates et al., 2003; Binder et al., 1997). One functional signature of these regions

is that they respond more when we process meaningful and structured linguistic representations (like sentences), compared to degraded linguistic stimuli (like lists of unconnected words or pseudo-words; e.g. Fedorenko, Hsieh, Nieto-Castañón, Whitefield-Gabrieli, & Kanwisher, 2010). Are these regions also sensitive to meaning and/ or structure above the sentence level? It does not appear so. Instead, brain imaging studies that have compared the processing of coherent texts vs. lists of unrelated sentences have reported activations in brain regions outside of the core fronto-temporal language network (e.g. Fedorenko & Thompson-Schill, 2014), including in the medial prefrontal cortex (MPFC), precuneus (PC), bilateral temporoparietal junction (TPJ), posterior superior temporal sulcus (PSTS), and anterior temporal lobes/poles, as well as some subcortical regions such as the hippocampus and the amygdala (e.g. Ferstl & von-Cramon, 2001; Ferstl, Neumann, Bogler, & von-Cramon, 2008; Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Maguire, Frith, & Morris, 1999; Mar, 2011; Xu, Kemeny, Park, Frattali, & Braun, 2005; Yarkoni, Speer, & Zacks, 2008).

The precise role of these regions in discourse comprehension, however, remains poorly understood. One hypothesis stems from the observation that these regions resemble a network of regions implicated in social cognition, including Theory of Mind (ToM), or the

---

ability to represent others' thoughts, beliefs, and desires (e.g. Adolphs, 2009; Mar, 2011; Saxe & Kanwisher, 2003). These ToM regions respond to diverse stimuli, both verbal and non-verbal, that evoke thoughts about others' mental states (e.g. Gallagher et al., 2000; see Mar, 2011; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014 for reviews) and appear to be content-specific (e.g. Saxe & Powell, 2006). One possibility is therefore that these regions are active during text comprehension because typical narratives, used in most prior studies involve animate entities, and understanding the narrative requires understanding the characters' mental states (Fletcher et al., 1995; Gallagher et al., 2000) and/ or those of the narrator.

An alternative hypothesis was advanced by Lerner, Honey, Silbert, and Hasson (2011), who examined the processing of a naturalistic linguistic narrative as well as versions of this narrative scrambled at different grain levels (paragraphs, sentences, or words). In an inter-subject correlation analysis (e.g. Hasson, Yang, Vallines, Heeger, & Rubin, 2008), they observed that regions anatomically similar to regions previously implicated in Theory of Mind show stronger inter-subject synchronisation for coherent texts compared to other conditions. They argued that ToM regions may be engaged during text comprehension because they integrate information over longer temporal windows, compared to the regions of the core language network (see also Blank, 2016; Blank & Fedorenko, in press).

To distinguish between these hypotheses, it is critical to examine the processing of expository texts, which typically lack social content / mental state attribution (Moss & Schunn, 2015; Moss, Schunn, Schneider, McNamara, & VanLehn, 2011; Swett et al., 2013). If activation within the Theory of Mind network is due to the ToM-rich content of typical narratives, then expository texts should not elicit activation in these regions. If, on the other hand, activation in the ToM network has to do with some content-independent processes related to the construction of discourse-level representations (e.g. integrating information over long temporal windows, as Lerner et al. have proposed), then we should find sensitivity to coherence even for expository texts.

Several prior studies have examined the processing of expository texts, but only a few were designed to isolate the cognitive processes specific to discourse-level comprehension (i.e. coherence building,e.g. Ferstl & von Cramon, 2002; Fletcher et al., 1995). For example, Ferstl and von Cramon (2002) investigated the role of the ToM network in discourse processing using a design that crossed coherence (coherent vs. unconnected pairs of sentences) and content (ToM vs. no ToM). They found activation in the MPFC, as well as the PC / posterior

cingulate cortex, for the coherent > unconnected contrast, even for no-ToM materials (although they also found that ToM materials elicited activation in the MPFC regardless of their coherence status). Based on their results, Ferstl & von Cramon argued against the content-specific hypothesis. Instead, they suggested that the MPFC supports "the initiation and maintenance of nonautomatic cognitive processes", which they argued are required for both ToM reasoning and coherence building.

Part of the difficulty in drawing clear conclusions from past studies has to do with their reliance on traditional group-level fMRI analyses, where the observed activations are interpreted with respect to coarse-level anatomy, via reverse inference (Poldrack, 2006; 2011). This approach is problematic given that functionally distinct regions often lie in close proximity to one another within the same macroanatomical area (e.g. Deen, Koldewyn, Kanwisher, & Saxe, 2015; Fedorenko, Duncan, & Kanwisher, 2012; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009). Thus, observing activation for discourse-processing manipulations within the MPFC – where activations for ToM tasks have been previously reported – does not warrant the conclusion that discourse processing and Theory of Mind share cognitive and neural resources (cf. Ferstl & von Cramon, 2002, Figure 1). This is especially important given the structural (e.g. Palomero-Gallagher, Zilles, Schleicher, & Vogt, 2013; Vogt, 2016) and functional (e.g. de la Vega, Chang, Banich, Wager, & Yarkoni, 2016) heterogeneity of the MPFC.

To shed further light on the role of ToM regions in discourse comprehension, we adopted a different approach. Using a ToM "localiser" task (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011; Saxe & Kanwisher, 2003), we first identified functional regions of interest (fROIs) – sets of voxels that are robustly engaged during ToM processing – in each individual participant. This localiser has been extensively used and shown to be robust across materials, modalities, and tasks (Jacoby, Bruneau, Koster-Hale, & Saxe, 2016; Skerry & Saxe, 2014). We then examined these fROIs' responses to critical conditions of interest: short coherent texts and sets of unrelated sentences, which require all the same linguistic processes (lexical, syntactic, and combinatorial semantic processing) except for inter-clause/ inter-sentence coherence building. Critically, to distinguish between the content-specific and content-independent hypotheses about the role of the ToM regions in discourse comprehension, we used expository texts, like the one at the beginning of the paper, with no animate entities and thus no mental states invoked. Thus, if (any of the) ToM regions respond more to texts than

unconnected sentences for these materials, that would suggest that those regions support discourse-level understanding in a content-independent fashion.

We had considered using a design similar to that used by Ferstl and von Cramon (2002). However, we opted for a simpler, two-condition manipulation because most of the comparisons enabled by the full 2×2 design are inherently confounded and thus difficult to interpret, or bear no relevance on the research question (about the mechanisms of coherence building). For example, the fundamental differences between *expository and narrative texts* make it impossible to match the materials for all the relevant (linguistic and other) features above and beyond the feature of interest (i.e. different content). In particular, narrative texts tend to be more accessible, easier to process and result in a deeper level of processing and hence greater memorability, are more relatable, etc. (see McNamara & Magliano, 2009 for a review). For some of these features, no agreed upon metrics even exist. To the extent that previous studies have compared coherent narratives varying in mental state content (e.g. Nijhof & Willems, 2015; Rice & Redcay,2016; Saxe & Powell, 2006; Tamir, Bricker, Dodell-Feder, & Mitchell, 2015), they have consistently reported stronger activity in the ToM network in the presence of mental state content (we leverage this property to localise the network). When comparing *narrative (rich in mental content) coherent vs. unconnected (non-coherent) materials*, we are confounding the two processes of interest: both the content-specific and content-independent hypotheses predict greater activity in the ToM regions for the coherent condition. The prediction of the content-independent hypothesis is clear: the coherent, but not the unconnected condition requires discourse-level structure building. The content-specific hypothesis makes the same prediction because the coherent condition would likely lead to richer, more elaborate representations of the characters and their mental states (Young, Dodell-Feder, & Saxe, 2010). Finally, the comparison between the two *unconnected conditions (expository vs. narrative)* is not of direct interest when studying coherence building. As a result, the comparison that stands a chance to dissociate between the content-independent and content-specific interpretations of the ToM network's engagement during discourse-level comprehension is that between coherent and unconnected materials that are free of mental state attribution (i.e. expository texts).

In addition to the ToM localiser, we included two other localiser tasks: i) the localiser for high-level language processing regions (Fedorenko et al., 2010),

and ii) the localiser for the domain-general multiple demand (MD) system (Duncan, 2010; 2013; Fedorenko, Duncan, & Kanwisher, 2013). The goal of including these localisers is three-fold. *First*, some studies have reported sensitivity to discourse-level processing within what-appear-to-be the regions of the core language network (e.g. Ferstl et al., 2008). However, the reverse-inference-based interpretation difficulty mentioned above applies here as well. Identifying core language-responsive regions at the individual-participant level will allow us to conclusively determine whether these regions are sensitive to meaning and structure above the sentence level. Furthermore, the posterior-most extent of the language activity in the parietal cortex falls within the angular gyrus, thus landing in close proximity to the temporoparietal junction. It is important to determine whether the previously reported responses to text coherence occur within the language vs. the ToM regions within this general anatomical area (e.g. Deen et al., 2015). *Second*, the language localiser will allow us to also define the right-hemisphere homologues of the language regions. These regions are important to examine because they have been linked to discourse processing based on the neuropsychological patient literature (e.g. Beeman, 1993; Sabbagh, 1999). And *third*, some have proposed that executive functions (like working memory and cognitive control) play some role in discourse-level comprehension (e.g. Barbey, Colom, & Grafman, 2013). Examining the responses of domain-general MD regions to the coherence manipulation will help evaluate this hypothesis.

Methods

Materials and task

Twenty-four texts were created based on Wikipedia (www.wikipedia.com) entries. Topics were selected so as to minimise mental content, avoiding articles that discuss particular individuals or social constructs (for the list of topics, see Table 1; the complete set of materials is available from https://osf.io/d86nc/). For each topic, we created a four-sentence-long text (see Table 1 for word count statistics). The texts were split into two groups of 12 for the purpose of the coherence manipulation, as discussed below. The groups were matched for average sentence length in each sentence position. In addition, each sentence had to satisfy the following criteria:

(1) describes factual information, with no evidence of mental states (either descriptions of mental states or attitudes of the narrator);

Table 1. Stimulus info.

		Stimuli Group 1						Stimuli Group 2			
Text #	Topic	S1 n words	S2 n words	S3 n words	S4 n words	Text #	Topic	S1 n words	S2 n words	S3 n words	S4 n words
1	Cheese	26	20	16	22	13	Pomelo	24	11	19	17
2	Lake Baikal	15	21	15	23	14	Ocean Current	20	20	21	19
3	Ice house	20	24	21	24	15	Refridgeration	14	19	23	24
4	Airplane	19	17	19	16	16	Vacuum	18	21	20	18
5	American Bison	17	22	17	17	17	Protein	13	16	19	21
6	New York City	21	26	22	23	18	Seattle	19	22	16	27
7	Tooth pick	12	13	14	22	19	Flag	24	23	25	23
8	Graphic design	21	17	14	16	20	Animation	23	18	17	14
9	Research	23	14	22	23	21	Knowledge	28	20	19	18
10	Porch	17	15	20	21	22	Floor (building)	18	19	15	21
11	Logging	25	26	23	15	23	Air	21	27	19	19
12	Art exhibition	14	24	14	20	24	School	16	11	11	15
	Mean	19.2	19.9	18.1	20.2		Mean	19.8	18.9	18.7	19.7
	Std.	4.3	4.6	3.5	3.3		Std.	4.4	4.6	3.7	3.7
	Min	12	13	14	15		Min	13	11	11	14
	Max	26	26	23	24		Max	28	27	25	27

Note: Information on the topics used and the number of words for each sentence.

(2) contains a sufficient number of "keywords", so that it is clear that it relates to the target topic and to the other sentences within the same topic;

(3) cannot be mistaken for a part of a text on a different topic within its subset of 12 texts, although some overlap in keywords among texts was allowed.

Any given participant saw either the texts in group 1 in the *Coherent* (*C*) condition and the texts in group 2 in the unconnected or *Non-Coherent* (*NC*) condition, or the texts in group 1 in the *Non-Coherent* condition and the texts in group 2 in the *Coherent* condition. In the *Coherent* condition, the four sentences from a given topic were presented in the correct order. In the *Non-Coherent* condition, four sentences from four different topics were presented, ensuring that no additional inferences are possible. However, each sentence appeared in the same position within the trial (first, second, third, or fourth) as the position it occupies in its original text. This preservation of sentence position between conditions ensured that within a trial (i.e. a set of four sentences), activation for a specific sentence was always at the same part of the trial. The scrambling was done separately for each participant. Sample stimuli are shown in Table 2.

A behavioural pre-test: a self-paced reading study

To ensure that our coherence manipulation was effective, we conducted a self-paced reading study using Mechanical Turk, Amazon.com's platform for collecting behavioural data. This is especially important given that we chose to use a passive reading task in the fMRI study, as described below (cf. studies that confound stimulus manipulations of coherence with a response bias by asking participants to make coherence judgments (e.g. Ferstl & von Cramon, 2002; see Egidi & Caramazza, 2016 for implications)).

As expected from ample previous research, coherent texts should be easier (faster) to process than sequences of unrelated sentences given that a) words are easier to process in supportive contexts because they become more predicable (e.g. Levy, 2008; Marslen-Wilson & Tyler, 1980; Smith & Levy, 2013), and b) in the *Coherent* condition, the supportive context keeps building up, with strong links among sentences, but in the *Non-Coherent* condition, there is coherence only within each sentence, but not across them (Graesser, Millis, & Zwaan, 1997; Haberlandt, 1980; Haberlandt & Graesser, 1985; Keenan, Baillet, & Brown, 1984; Kintsch, Mandel, & Koz-

Table 2. Sample stimulus.

	Coherent condition	Non-coherent condition
First sentence	A pomelo is usually pale green to yellow when ripe, with sweet white (or, more rarely, pink or red) flesh and very thick rind.	A pomelo is usually pale green to yellow when ripe, with sweet white (or, more rarely, pink or red) flesh and very thick rind.
Second sentence	It is a large citrus fruit, usually weighing around 3 pounds.	Labels on packets of cheese often claim that a cheese should be consumed within three to five days of opening.
Third sentence	The fruit tastes like a sweet, mild grapefruit, and has none, or very little, of the common grapefruit's bitterness.	Lake Baikal was formed as an ancient rift valley, having the typical long crescent shape.
Fourth sentence	The peel is sometimes used to make marmalade, can be candied, and is sometimes dipped in chocolate	This circulation plays an important role in supplying heat to the Polar Regions, and thus in sea ice regulation.

Note: Sample stimulus from the coherent and non-coherent conditions

minsky, 1977; Radach, Huestegge, & Reilly, 2008; Wochna & Juhasz, 2013).

Methods

We posted surveys for 100 workers on Mechanical Turk using the self-paced reading software developed by Hal Tily (e.g. Singh, Fedorenko, Mahowald, & Gibson, 2016). All workers were paid for their participation. Participants were asked to indicate their native language, but payment was not contingent on their responses.

The materials were divided into two lists, as described above, and any given participant saw one list. The only difference from the fMRI study was that instead of scrambling the sentences separately for each participant to create the *Non-Coherent* condition, the materials were scrambled once within each list (with a constraint that no two sentences from the same topic appear in the same text), and the same scrambled versions were used for all participants within a list. The order of trials within each list was randomised for each participant using the Turkolyzer software (Gibson, Piantadosi, & Fedorenko, 2011). To ensure that participants read the materials for meaning, simple yes/no comprehension questions were included after each text. The question always referred to the information that was provided in the third sentence of each stimulus.

Each trial began with a prompt screen. Pressing the space bar revealed each consecutive sentence or part of the sentence (sentences that were longer than 80 characters with spaces were divided into two parts, to avoid line-wrapping). After the last sentence or part of the sentence, the comprehension question appeared, and participants had to press one of two buttons to respond. Participants were told whether or not they answered correctly. The response time (RT) was recorded for every button press.

Results

Ninety-four out of 100 participants indicated English as their native language (48 for list 1 and 46 for list 2). We excluded those who did not indicate so, along with four additional participants with overall comprehension accuracies below 0.65; for the remaining participants, the average accuracy was 0.835. This left 90 participants for reading time (RT) analyses (48 for list 1 and 42 for list 2). Next, we excluded each sentence for which the RT was shorter than 1 s, or sentence part for which the RT was shorter than 0.3 s or longer than 8 s. For any given trial, if the RT for a part of a sentence satisfied the criteria above, we removed the whole sentence from the analyses. We then calculated RTs for each sentence (summing across sentence parts where applicable). Due to a scripting error, 12 of the 96 sentences (each from

a different text) were divided into parts differently across the two lists, making the comparisons for those sentences difficult. We therefore excluded those sentences from the analysis. As a result of the data exclusions, we were left with 6,730 observations (70%) of single sentence RTs (out of the original planned 9,600).

Reading time analyses were conducted using lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for the statistical language R (R core Team,2013). First, we made sure that there was no effect of condition on the RT for the first sentence (when participants could not know which condition the trial belonged to). To do so, we fitted a mixed-effects model, predicting the RT of the first sentence from condition with random intercepts and slopes by condition for participants and items (Barr, Levy, Scheepers, & Tily, 2013). As expected, there was no evidence for the effect of condition ($p = .33$).

To test for the effect of condition on the sentences in the second, third and fourth positions, we fitted a similar maximal mixed-effects model. We treated each sentence as an item (resulting in 65 unique items, out of the planned 72, due to exclusions). There was a strong effect of condition, such that sentences were read faster in the *Coherent* condition ($p < .001$; see Figure 1).

Finally, we tested for the effect of condition on comprehension accuracies using a logistic regression, and found no effect ($p = .091$). Accuracies were high in both conditions: 85.8% and 82.8% in the *Coherent* and *Non-Coherent* conditions, respectively.

The data and the R analysis files are available at https://osf.io/d86nc/.

Participants

Seventeen right-handed adults (10 females, 7 males, mean age 23.7, range 19-34) participated in the study for payment. All participants were native English
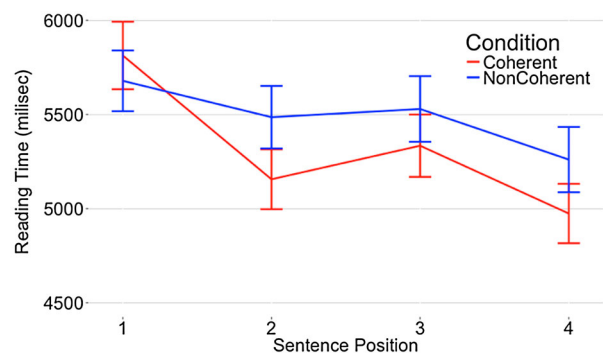

Figure 1. Reading times in the self-paced reading study. Reading times are averaged across items for each sentence position in each condition. Error bars represent standard errors of the mean by participants.

speakers and had normal or corrected-to-normal vision. All participants gave written informed consent in accordance with the requirement of MIT's Committee on the Use of Humans as Experimental Subjects.

Each participant completed three localisers and the critical text comprehension task. Some participants performed additional unrelated experiments in the same session. Some of the participants had completed one or more of the localisers during earlier scanning sessions.

Two of the participants were excluded due to excessive motion and falling asleep in the scanner, leaving fifteen participants (10 females, mean age 23.7, range 19-34) for analysis. This is a reasonable sample size for experiments that use an individual functional regions of interest approach (Fedorenko et al., 2010; Nieto-Casta-ñón & Fedorenko, 2012).

Functional localisers

Tom localiser

Participants read stories in a slow event-related design. In the *False Belief* condition, an outdated representation was held by a person; in the *False Photo* condition, an outdated representation was contained in an inanimate object, like a picture or a map. The *False Belief > False Photo* contrast targets brain regions engaged in Theory of Mind processing (Dodell-Feder et al., 2011; Jacoby et al., 2016; Saxe & Kanwisher, 2003).

Each trial started with story presentation for 10 s, followed by 4 s to answer a question about the story, for a total trial duration of 14 s. Each run consisted of 10 trials (5 per condition) and 11 fixation blocks of 12 s, for a total duration of 272 s (4 min 32 s). Each participant performed two runs. Condition order was counterbalanced across runs.

Language localiser

Participants read sentences and lists of pronounceable nonwords presented one word/nonword at a time in a blocked design. The *Sentences > Nonwords* contrast targets brain regions sensitive to high-level linguistic processing (Fedorenko et al., 2010; Fedorenko, Behr, & Kanwisher, 2011).

Each trial started with 100 ms pre-trial fixation, followed by a 12-word-long sentence or a list of 12 nonwords presented on the screen one word/nonword at a time at the rate of 450 ms per word/nonword. Then, a line drawing of a finger pressing a button appeared for 400 ms, and participants were instructed to press a button whenever they saw this icon, and finally a blank screen was shown for 100 ms, for a total trial duration of 6 s. The button-pressing task was included to help participants stay awake and focused. Each block consisted of

3 trials and lasted 18 s. Each run consisted of 16 experimental blocks (8 per condition), and 5 fixation blocks of 14 s, for a total duration of 358 s (5 min 58 s). Each participant performed two runs. Condition order was counterbalanced across runs.

MD localiser

Participants performed a spatial working memory task that has been previously shown to activate the MD system broadly and robustly (Fedorenko et al., 2013). Subjects had to keep track of four (*Easy* condition) or eight (*Hard* condition) locations in a 3×4 grid (Fedorenko et al., 2011). In both conditions, subjects performed a two-alternative forced-choice task at the end of each trial to indicate the set of locations that they just saw. The *Hard > Easy* contrast targets brain regions engaged in cognitively demanding tasks. Fedorenko et al. (2013) have shown that the regions activated by this task are also activated by a wide range of other executive function tasks contrasting a difficult vs. an easier condition.

Each trial lasted 8 s (see Fedorenko et al., 2011 for details). Each block consisted of 4 trials and lasted 32 s. Each run consisted of 12 experimental blocks (6 per condition), and 4 fixation blocks of 16 s, for a total duration of 448 s (7 min 28 s). Each participant performed one or two runs. Condition order was counterbalanced across runs when participants performed two runs.

Critical task

Participants read the stimuli from the *Coherent* and *Non-Coherent* conditions presented one sentence at a time. To ensure that participants paid attention, they were instructed that they would perform a memory task in the scanner following the reading scans.

Any given participant saw 12 stimuli per condition (list 1 or list 2), distributed across two runs. After the two runs, participants completed a memory test where they were presented with a set of 21 sentences and asked – for each sentence – whether they had encountered it during the experiment. The set contained 12 sentences from the experimental materials (6 from each condition), always taken from the third position, and 9 novel sentences.

Each trial lasted 24 s and consisted of 4 sentences each presented for 6 s. Each run consisted of 12 trials (6 per condition) interleaved with fixation periods. There were 5 fixation periods of 14 s: at the beginning and end of the run, and after each fourth experimental trial. Other trials were separated by fixation periods of 4 s. Each run thus lasted 380 s (6 min 20 s). Condition order was counterbalanced across runs and participants.

After participants performed the memory task, they completed two additional runs where they saw the stimuli from the other list (i.e. list 2 if they first saw list 1). These data were not analysed for the purposes of the current study because we wanted to make sure that in the *Non-Coherent* condition participants wouldn't be retrieving the broader contexts associated with familiar sentences presented in the *Coherent* condition during the first two runs.

fMRI data acquisition

Participants were scanned using a Siemens Magnetom Tim Trio 3 T system (Siemens Solutions, Erlangen, Germany) in the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT with a 32-channel head coil. First, high-resolution structural (anatomical) images were acquired using T1MPRAGE sequence in 128 axial slices with 1.33 mm isotropic voxels (TE = 3.39 ms). All functional scans used a full brain coverage sequence, acquiring 31 4 mm thick near-axial 96 × 96 slices in interleaved order (voxel size of 2.083 × 2.083 × 4 mm, TR = 2 s, TE = 30 ms, flip angle = 90°). The first 10 s of each functional run were excluded to allow for steady state magnetisation. While participants were performing the memory test after the first two runs of the critical task, a diffusion tensor imaging sequence was acquired for an unrelated study.

Data analysis

Functional data were preprocessed and modelled using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/), and custom software. The data were motion corrected and co-registered to the first scan of each functional run using rigid body transformations. All functional runs were co-registered to one another and then to the structural image. The structural image was co-registered and normalised to Montreal Neurological Institute (MNI) brain template, using SPM's non-linear warp algorithm. The normalising transformation was then applied to the co-registered functional data to bring all the data to a common brain space. Lastly, all functional data were smoothed using a 5 mm FWHM Gaussian kernel filter.

All first-level analyses – for both the localisers and the critical task – were performed by fitting a general linear model (GLM) to the relevant imaging data. Each model included condition regressors, all of which were block design boxcar functions, following the onsets and durations of the items presented. The boxcars were convolved with a canonical double gamma hemodynamic response function (HRF) to create the GLM regressors.

The models included covariates of no interest, accounting for intercept term and run effects. Time series were subjected to high pass filter (cut-off of 1/128 Hz).

Functional region of interest (fROI) definition

Data from the localiser tasks were modelled for each participant as described above. The appropriate contrast of interest was then used to create a whole-brain statistical parametric map of *t* statistics to identify voxels recruited by the relevant localiser task. For each functional network – ToM, language, and MD – we used a set of parcels or "search spaces" (i.e. brain areas within which most individuals in prior studies showed activity for the relevant localiser contrast), which were combined with each individual participant's activation map for the relevant contrast to define subject-specific fROIs (Fedorenko et al., 2010; Julian, Fedorenko, Webster, & Kanwisher, 2012).

To define the ToM fROIs, we used five parcels (Figure 2 (a)) derived from a group-level activation map for the *False Belief > False Photograph* contrast in 462 participants (Dufour et al., 2013). The parcels were created from a random effects analysis on the contrast of interest from all 462 subjects. The resulting *t*-map was thresholded for $t > 3$ ($p < .0014$) and minimum voxel extent of 10. Due to the robustness and extent of the resulting clusters, physical boundaries were imposed to separate some of the clusters. The resulting parcels included regions in medial prefrontal cortex (MPFC, encompassing both dorsal and middle parcels reported in Dufour et al. (2013) and restricted physically by z(mni) > 0), ventral medial prefrontal cortex (VMPFC, restricted by z(mni) < 0), precuneus (PC), right temporo-parietal junction (RTPJ, restricted by z(mni) > 4), and left temporo-parietal junction (LTPJ, restricted by z(mni) > 7). The ToM parcels are available for download from http://saxelab.mit.edu/ToMgroupMaps.php.

To define the language fROIs, we used six parcels (Figure 2(b)) derived from a group-level representation of data for the *Sentences > Nonwords* contrast in 220 participants following the group-constrained subject-specific (GSS) procedure described in Fedorenko et al. (2010). These parcels included three regions in the left frontal cortex: two located in the inferior frontal gyrus (LIFGorb and LIFG), and one located in the middle frontal gyrus (LMFG), and three regions in the left temporal and parietal cortices spanning the entire extent of the lateral temporal lobe and going posteriorly to the angular gyrus (LAntTemp, LPostTemp, and LAngG). These parcels are similar to the parcels reported originally in Fedorenko et al. (2010) based on a set of 25 participants, except that the two anterior temporal parcels

(LAntTemp, and LMidAntTemp) ended up being grouped together, and the two posterior temporal parcels (LMidPostTemp and LPostTemp) ended up being grouped together. We also defined the right-hemisphere homologue of the language network. To do so, the left-hemisphere parcels were mirror-projected onto the right hemisphere to create six homologous parcels, as in Blank, Kanwisher, and Fedorenko (2014).

The language, and the MD parcels (see below), are available for download from https://evlab.mit.edu/funcloc/download-parcels.

To define the MD fROIs, following Fedorenko et al. (2013), we used eighteen anatomical regions (Figure 2 (c)) bilaterally (Tzourio-Mazoyer et al., 2002) previously implicated in MD activity: opercular IFG (LIFGop & RIFGop), MFG (LMFG & RMFG), orbital MFG (LMFGorb & RMFGorb), precentral gyrus (LPrecG & RPrecG), insular cortex (LInsula & RInsula), supplementary and presupplementary motor areas (LSMA & RSMA), inferior parietal cortex (LParInf & RParInf), superior parietal cortex (LParSup & RParSup), and anterior cingulate cortex (LACC & RACC).

Within each ToM, language, and MD parcel, we selected the top 10% of most localiser-responsive voxels based on the *t*-values for the relevant contrast (*False Belief > False Photograph*, *Sentences > Nonwords*, and *Hard > Easy spatial WM*, respectively). This approach ensures that a fROI can be defined in every participant, and that the fROI sizes are identical across participants (Nieto-Castañón & Fedorenko, 2012).

To validate the response of the fROIs to the relevant localiser contrast, we used a cross-validation procedure. For each localiser and each ROI, we performed the selection procedure above separately for each of the two runs and extracted the beta estimates for the response to the localiser conditions from the left-out run. We then averaged the independently extracted responses across runs, and used a *t*-test to validate the sensitivity of the fROIs to the relevant localiser contrast.

To estimate the response of the fROIs to the conditions of the critical task, data from all the localiser runs were used to define the fROIs for each of the three functional networks. For each fROI, we then extracted the model's beta estimates for the two conditions (*Coherent*, *Non-Coherent*) in the critical task. These beta estimates were used for the inferential statistics investigating the effects of the experimental manipulation between and within the functional networks of interest.

To facilitate comparisons with prior studies, we additionally performed a whole-brain random-effects analysis for the critical coherence task. And in a complementary exploratory analysis, we characterised the main significant clusters with respect to their responses to the localiser tasks. In particular, we used the clusters as search spaces to define individual fROIs based on the *Coherent > Non-Coherent* contrast. We then extracted the responses of those fROIs to the localiser tasks. This latter analysis was performed to ensure that we have not missed any important coherence-sensitive regions that might fall outside of our three networks of interest (ToM, language, and MD). Because these analyses do not directly address our research question / test the critical hypotheses, we make the random-effects *t*-map and the results of the additional ROI analysis available (https://osf.io/d86nc/), but do not discuss the results below. (Briefly: the results were in line with the critical hypothesis-driven fROI-based analyses discussed below.)

Results

Results: memory test

Overall, participants performed well, with the average accuracy of 0.86. No significant difference was observed between the coherent and non-coherent conditions in either the accuracies ($\mu(C) = 0.8$; $\mu(NC) = 0.82$; $t(14) = 0.59$, *n.s.*) or reaction times ($\mu(C) = 3.81s$; $\mu(NC) = 3.61s$; $t(14) = 0.47$, *n.s.*). Participants were more accurate rejecting novel sentences than recognising familiar ones ($t(14) = 1.76$, $p = .049$), but showed no difference in reaction time ($t(14) = 0.43$, *n.s.*). These results confirm that participants paid attention and read the materials for meaning.

Results: localiser tasks

Replicating much prior work, the functional localisers all showed robust effects as assessed with the cross-validation procedure described above (Table 3). Note that for two of the participants we had only one functional run of the MD localiser task. Those participants were therefore excluded from the cross-validation analysis for the MD network, but we ensured that the MD localiser elicited the expected activation patterns by examining individual whole-brain maps.

Results: critical task

All the following analyses were done after extracting the beta estimates for the two conditions of the main task (*Coherent* and *Non-Coherent*) from each of the fROIs.

Networks comparison

The first analysis was a between-network omnibus one-way anova. To do this, we first averaged the response to each condition across the fROIs within each of the

Table 3. Localisers results.

Theory of mind network					Multiple demand network				
False belief > false photo					Spatial working memory hard > easy				
Region	Significant	t statistic	df	p value	Region	Significant	t statistic	df	p value
MPFC	TRUE	5.63	14	<.001	Left Hemisphere				
VMPFC	TRUE	4.47	14	.001	LIFGop	TRUE	3.60	11	.004
PC	TRUE	13.40	14	<.001	LMFG	TRUE	4.02	11	.002
RTPJ	TRUE	7.32	14	<.001	LMFGorb	TRUE	4.03	11	.002
LTPJ	TRUE	8.72	14	<.001	LPrecG	TRUE	3.37	11	.006
					LInsula	TRUE	3.66	11	.004
					LSMA	TRUE	4.20	11	.001
Language network					LParInf	TRUE	5.85	11	<.001
Sentences > nonwords					LParSup	TRUE	7.51	11	<.001
Region	significant	t statistic	df	p value	LACC	TRUE	2.78	11	.018
Left hemisphere					Right Hemisphere				
LIFGorb	TRUE	7.01	14	<.001	RIFGop	TRUE	4.53	11	.001
LIFG	TRUE	7.36	14	<.001	RMFG	TRUE	4.76	11	.001
LMFG	TRUE	8.14	14	<.001	RMFGorb	TRUE	5.79	11	<.001
LAntTemp	TRUE	7.95	14	<.001	RPrecG	TRUE	3.60	11	.004
LPostTemp	TRUE	8.45	14	<.001	RInsula	TRUE	3.44	11	.006
LAngG	TRUE	6.28	14	<.001	RSMA	TRUE	3.80	11	.003
Right hemisphere					RParInf	TRUE	4.72	11	.001
RIFGorb	TRUE	3.67	14	.003	RParSup	TRUE	6.70	11	<.001
RIFG	TRUE	4.51	14	<.001	RACC	TRUE	3.38	11	.006
RMFG	TRUE	3.95	14	.001					
RAntTemp	TRUE	6.85	14	<.001					
RPostTemp	TRUE	6.49	14	<.001					
RAngG	TRUE	4.84	14	<.001					

Note: Reliability of the localiser responses in each of the three functional networks. The comparisons are performed on beta values extracted from data independent from the data used to define the fROIs, as described in Methods.

five networks of interest (ToM, LH/RH language, LH/RH MD; NB: we divided the MD network into LH and RH components, to mirror the analysis for the language network; we did not do this for the ToM network given that it includes several cortical midline regions). We then ran the one-way (interaction) anova with condition and network as within-subject variables. The analysis revealed significant main effects of both condition ($F(1,14) = 24.99$, $p < .001$) and network ($F(4,56) = 41.47$, $p < .001$), and, importantly, a significant interaction between the two ($F(4,56) = 8.58$, $p < .001$). This pattern suggests that different networks show differential sensitivity to the critical task.

We then analysed the responses of each network in greater detail. In particular, within each network we examined the responses of each fROI to the two conditions using a t-test; the resulting significance values were corrected for multiple comparisons (i.e. number of ROIs in the relevant network). In addition, to estimate the effect of the critical manipulation on each of the networks, we fit a multilevel model (per network) with participants treated as random effects within the population, and fROIs – as random effects within the network. We then fit a second, reduced model with the same random effects structure but with no effect of condition. Finally, we compared those models to estimate the effect of the manipulation (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000).

Within network:

$$\beta = condition + (condition|subject) + (condition|fROI)$$

Theory of Mind network

The only regions in the ToM network that showed a reliable between-condition difference were the medial frontal fROIs. Both showed a stronger response for the *Coheren*t condition (MPFC: $t(14) = 3.55$, $p = .003$; VMPFC: $t(14) = 3.13$, $p = .007$). In addition, the LTPJ showed a statistical difference ($t(14) = 2.53$, $p = .024$) which did not survive the correction for multiple ($n = 5$) comparisons. The model comparison showed a marginal preference for the full model, indicating a significant effect of condition on ToM network ($\chi^2(1) = 3.47$, $p = .062$).

To further explore the observed differences among the ToM ROIs, we ran a one-way anova on the effect of coherence (difference between beta estimates for the *Coherent* and *Non-Coherent* conditions) across ROIs. The analysis showed a statistically significant main effect of ROI ($F(4,56) = 5.57$, $p < .001$). We then directly compared ROIs using post-hoc t-tests (here and in all subsequent analyses, two-tailed t-tests were used). Specifically, we compared RTPJ to MPFC ($t(14) = 3.93$, $p = .0015$), VMPFC ($t(14) = 2.16$, $p = .049$) and LTPJ ($t(14) = 2.68$, $p = .017$). These results show that the ROIs not only differ in whether or not they show a significant effect of coherence, but they are significantly different from one

another in their response patterns (e.g. see Nieuwenhuis, Forstmann, & Wagenmakers, 2011 for discussion). Thus, different regions of the ToM network are differentially engaged in the processing of coherence during text-level understanding.

Left hemisphere language network

In spite of overall strong responses to both conditions – to be expected given the linguistic nature of the materials – none of the LH language fROIs showed sensitivity to the coherence manipulation (all $ps > 0.1$ uncorrected). Correspondingly, the multi-level model comparison, across the LH language network did not show a significant effect of condition ($\chi^2(1) = 0.094$, $p = .759$).

Right hemisphere language network

The right hemisphere homologues of the LH language fROIs showed an overall lower response to the two conditions, compared to the LH language fROIs (Mahowald & Fedorenko, 2016). In addition, two of the frontal ROIs showed marginally stronger responses to the Non-Coherent condition, but these effects did not survive the correction for multiple ($n = 6$) comparisons (RIFG: $t(14) = 2.62$, $p = .02$; RIFGorb: $t(14) = 2.855$, $p = .013$). Similar to its LH counterpart, the RH language network model

comparison did not show a significant effect of condition ($\chi^2(1) = 0.546$, $p = .46$).

Left hemisphere multiple demand network

Most LH MD fROIs, except for the LPrecG and LParSup fROIs, showed stronger responses to the Non-Coherent condition, with several ROIs surviving the correction for multiple ($n = 9$) comparisons (see Table 4 for details). The multi-level model comparison across the LH MD network showed a significant effect of condition ($\chi^2(1) = 6.98$, $p = .008$).

Right hemisphere multiple demand network

All the RH MD fROIs showed significantly stronger responses to the Non-Coherent condition, with several ROIs surviving the multiple-comparisons correction (see Table 4 for details). As in the LH MD network, the multi-level model comparison across the RH MD network showed a significant effect of condition ($\chi^2(1) = 11.815$, $p < .001$).

Discussion

Prior studies that investigated the processing of coherence in discourse (e.g. Ferstl & von-Cramon, 2001, 2002; Kuperberg et al., 2006; Maguire et al., 1999; Xu et al., 2005) have reported activation in brain regions that

Table 4. Critical task results.

Theory of mind network					Multiple demand network				
Corrected criteria: $p < .01$					Corrected criteria: $p < .0055$				
Region	significant	t statistic	df	p value	Region	significant	t statistic	df	p value
MPFC	TRUE	3.55	14	.003	Left hemisphere				
VMPFC	TRUE	3.13	14	.007	LIFGop	UNCORRECTED	−2.32	14	.036
PC	FALSE	0.73	14	.480	LMFG	TRUE	−3.58	14	.003
RTPJ	FALSE	−0.13	14	.899	LMFGorb	UNCORRECTED	−2.26	14	.041
LTPJ	UNCORRECTED	2.53	14	.024	LPrecG	FALSE	−1.75	14	.101
					LInsula	TRUE	−4.92	14	<.001
Language network					LSMA	TRUE	−3.78	14	.002
Corrected criteria: $p < 0.0083$					LParInf	UNCORRECTED	−2.52	14	.024
Region	significant	t statistic	df	p value	LParSup	FALSE	−1.58	14	.137
Left Hemisphere					LACC	UNCORRECTED	−3.01	14	.009
LIFGorb	FALSE	−1.56	14	.141	Right hemisphere				
LIFG	FALSE	−0.90	14	.383	RIFGop	TRUE	−3.75	14	.002
LMFG	FALSE	−1.28	14	.221	RMFG	TRUE	−3.68	14	.002
LAntTemp	FALSE	−0.56	14	.586	RMFGorb	TRUE	−3.35	14	.005
LPostTemp	FALSE	0.52	14	.609	RPrecG	UNCORRECTED	−2.18	14	.047
LAngG	FALSE	1.67	14	.117	RInsula	TRUE	−3.47	14	.004
Right Hemisphere					RSMA	TRUE	−4.12	14	.001
RIFGorb	UNCORRECTED	−2.86	14	.013	RParInf	UNCORRECTED	−2.85	14	.013
RIFG	UNCORRECTED	−2.62	14	.020	RParSup	UNCORRECTED	−2.81	14	.014
RMFG	FALSE	−1.76	14	.100	RACC	TRUE	−3.94	14	.001
RAntTemp	FALSE	0.23	14	.823					
RPostTemp	FALSE	1.18	14	.259					
RAngG	FALSE	1.91	14	.077					

Note: Region of interest analyses by network. We report the results of two-tailed t-tests examining the Coherent > Non-Coherent effect in the ToM, language, and MD fROIs defined in individual participants, as described in Methods. Significance values are corrected for the number of fROIs in each network (ToM: $n = 5$; LH/RH language: $n = 6$; LH/RH MD: $n = 9$); if an effect does not survive the multiple-comparisons correction, it is marked as "uncorrected" in the "significant" column.
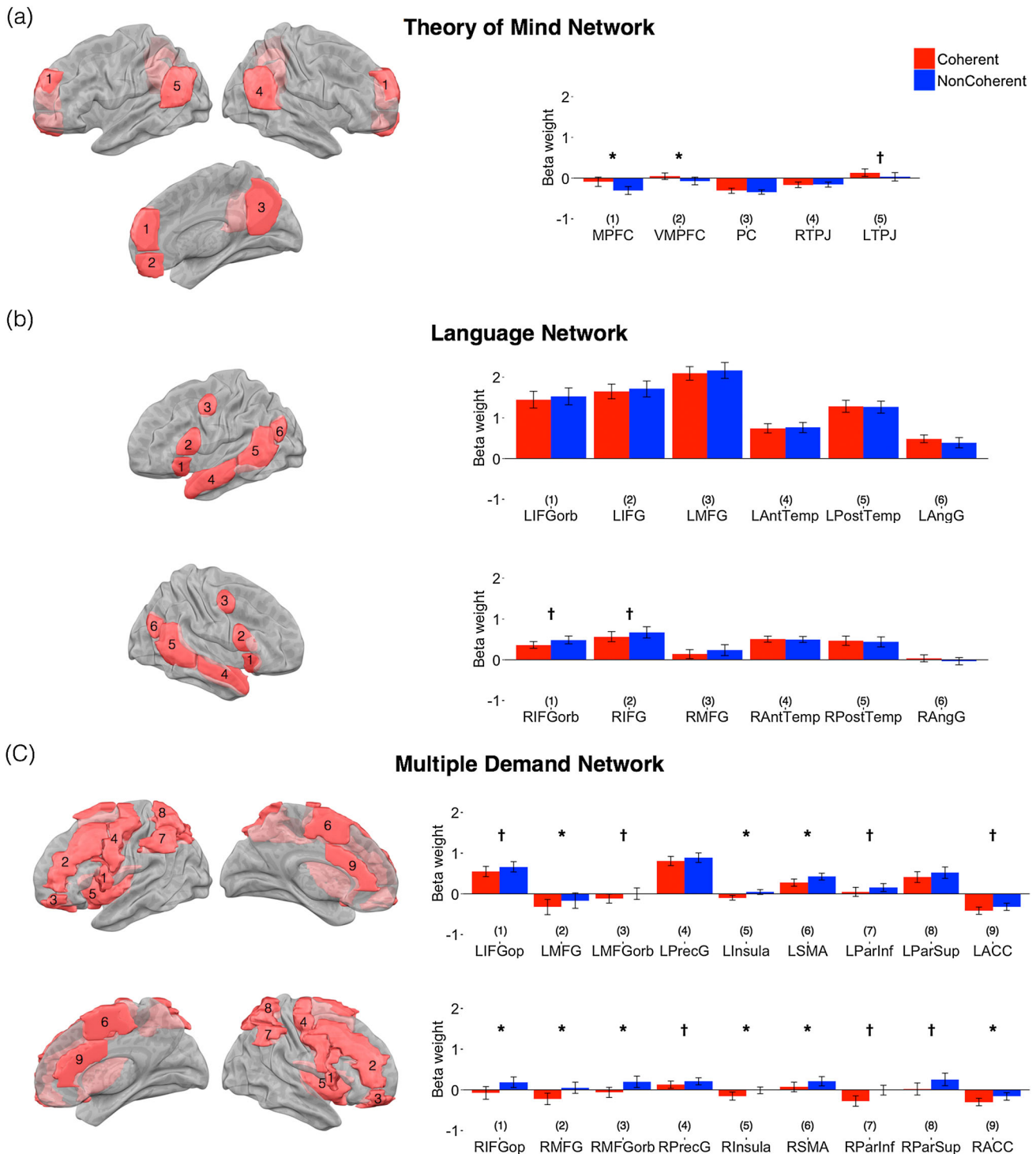
Figure 2. Responses of all fROIs to the coherence manipulation: (a) ToM network; (b) Language network (top: left hemisphere, bottom: right hemisphere); (c) Multiple Demand network (top: LH, bottom: RH). For each network, on the left we show the search parcels used to constrain the definition of individual fROIs (see Methods; note that the individual fROIs constitute 10% of each parcel). Error bars represent standard errors of the mean by participants. Statistical significance for the within participant t-test per fROI is marked with a * ($p < .05$ for the Bonferroni-corrected level by network) or with a † ($p < .05$ uncorrected).

resemble the Theory of Mind network (e.g. Ferstl et al., 2008; Mar, 2011; Saxe & Kanwisher, 2003). Two broad hypotheses have been put forward about the role of these brain regions in discourse-level comprehension.

One hypothesis has to do with the *content* of typical narratives. According to this hypothesis, ToM regions are active during text processing because most texts require thinking about the characters' mental states and/or

those of the narrator (e.g. Gallagher et al., 2000; Saxe & Powell, 2006), and texts allow for more and/or richer mental state inferences than unconnected sentences.

According to the alternative hypothesis, the responses to text coherence in these brain regions are *content-independent*. For example, Lerner et al. (2011; see also Blank & Fedorenko, in prep.) have argued that ToM regions are engaged in text processing because they integrate information over longer temporal windows. And Ferstl and von Cramon (2002) have argued that text-level comprehension and thinking about others' mental states share domain-general computational demands having to do with "non-automatic cognitive processes", such as general inferencing.

To distinguish between the content-dependent and content-independent hypotheses, we investigated the processing of expository texts that do not involve animate entities and do not discuss the narrator's attitudes thus minimising mentalising demands. To circumvent the problem of reverse inference (Poldrack, 2006, 2011), we adopted an approach where we functionally localised ToM regions in each participant individually and then examined their responses to the two critical conditions: coherent texts or sequences of unconnected sentences. To additionally evaluate other ideas from the literature – about the role of linguistic and executive processes in discourse-level comprehension (as discussed in the Introduction) – we defined four other sets of brain regions: left-hemisphere language processing regions (e.g. Fedorenko et al., 2010), their right-hemisphere homologues, and LH and RH regions of the fronto-parietal multiple demand (MD) network (e.g. Duncan, 2010; 2013).

The results revealed strong responses to both the *Coherent* and the *Non-Coherent* conditions in the LH language regions, to be expected given that both conditions require lexical and combinatorial syntactic/ semantic processing (Fedorenko et al., 2010; Fedorenko, Nieto-Castañón, & Kanwisher, 2012). However, the two conditions did not differ in the mean level of response. A similar picture obtained in the RH homologues of the language regions: the overall responses were lower than in the LH, but as the LH regions, most RH regions did not differentiate between the two conditions. Two regions (the RIFGorb and RMFG fROIs) responded more strongly during the processing of the *Non-Coherent* condition. These results pose a challenge for the claims that RH language regions support the processing of coherent narratives and/or non-literal processing (e.g. inferencing) more generally (e.g. Beeman, 1993; Beeman, Bowden, & Gernsbacher, 2000; Just & Varma, 2007).

The domain-general MD network responded more strongly to the *Non-Coherent* condition bilaterally, with many regions showing this effect reliably. Given that we have observed greater processing difficulty in the *Non-Coherent* condition in our behavioural data, the MD response is in line with the general sensitivity of this network to effort (Duncan & Owen, 2000; Fedorenko et al., 2013). The greater difficulty in the *Non-Coherent* condition could result from the greater number of mental switches among the different topics, and/or from the violated expectations about the later sentences continuing the topic introduced in the first sentence.

Critically, some regions of the ToM network responded reliably more strongly during the processing of coherent narratives than sequences of unconnected sentences. In particular, this pattern was observed in the medial frontal ToM fROIs, and marginally in the LTPJ fROI. Posterior ToM regions (the RTPJ and PC fROIs) did not differentiate between the two conditions in the mean level of response. These differences between the medial frontal and posterior ToM fROIs were significant, providing evidence for distinct contributions of different ToM regions to text comprehension / coherence processing.

Before discussing the implications of these results, it is worth noting that the effect of coherence was small in size. This is surprising given the richness and complexity of the representation that emerges when we process connected texts relative to unconnected sentences (e.g. Hobbs, 1985; Kintsch, 1998; Wolf & Gibson, 2005). It is also striking given the much larger effect size – localised to the fronto-temporal language network – of the sentences vs. word list contrast, where sentences elicit a response approximately twice the size of that elicited by unconnected words (e.g. Fedorenko et al., 2010; Pallier, Devauchelle, & Dehaene, 2011; Snijders et al., 2009). Thus, it appears that within-sentence syntactic and semantic composition are not only carried out by distinct mechanisms from those that support the construction of discourse-level representations, but also require substantially greater resources. The whole-brain analysis that we conducted ensured that we didn't miss a region (or regions) outside of our brain networks of interest that shows a large effect of coherence. The small size of the coherence effect should be kept in mind when theorising about the functions of the medial ToM regions and about coherence building in general. One possibility to be evaluated in future work is that actively encouraging (particular kinds of) inferences may lead to greater responses to coherent texts.

The small size of the coherence effect aside, our results have a number of implications. *First*, the lack of a stronger response to coherent texts than sentence sequences in posterior ToM fROIs in the current study – in light of prior coherent > non-coherent effects reported

throughout the ToM system (e.g. Kuperberg et al., 2006; Xu et al., 2005; Yarkoni et al., 2008) – suggests that those regions are plausibly content-specific and only get engaged when the stimuli require consideration of others' mental states (Saxe & Powell, 2006). In line with our findings, Ferstl and von Cramon (2002), as well as Lin et al. (2018) in a recent study, have found that parts of the MPFC (and the LTPJ) support discourse processing regardless of content, whereas the RTPJ only engages in coherence building for ToM materials. Both of these studies have, however, relied on traditional group analyses, which do not allow strong inferences about the coherence effects actually originating within ToM-responsive cortex (cf. nearby within the broader anatomical areas). We circumvented this problem with functionally defining ToM regions and provided clear evidence of distinct contributions of medial frontal vs. posterior ToM regions to coherence processing in texts, with only the former engaging even when the texts have no/minimal mental state content. Relatedly, the engagement of RTPJ and PC during narrative processing in Lerner et al.'s (2011) study is plausibly driven by greater mentalising demands of paragraphs and texts compared to single sentences, rather than by the fact that information has to be integrated over longer windows.

Second, responses to coherence in the medial frontal ToM fROIs suggest that these parts of the ToM network are not content-specific. Instead, they plausibly support some computation(s) required for building a discourse-level representation. The nature of these computations remains to be determined. At present, the available evidence is consistent with both the longer information integration window idea (Lerner et al., 2011) and general inferencing processes (e.g. Ferstl & von Cramon, 2002; see also, Ferstl et al., 2008; Friese, Rutschmann, Raabe, & Schmalhofer, 2008; Kuperberg et al., 2006). However, one prior hypothesis does not appear likely given our results. In particular, Mason and Just (2009, 2011) have proposed that medial frontal regions support the updating of situation models (Graesser et al., 1997; Van Dijk & Kintsch, 1983) via protagonist monitoring. The materials used in the current study do not allow for the construction of a situation model around a protagonist or, more generally, for meaningful sequencing of events and inferences about causal relations between events. However, even though not content-specific, MPFC does exhibit stronger responses to social/mental content over and above the effect of coherence (Dodell-Feder et al., 2011; Lin et al., 2018). This preference should be taken into account in further theorising about the contributions of MPFC to discourse-level comprehension and about the mental processes that underlie coherence building. Comparing

responses to different kinds of coherence relations (e.g. Wolf & Gibson, 2005) may help constrain the possibilities.

Third, the fact that language regions failed to differentiate between coherent narratives and unconnected sentences – in spite of strong overall response to both conditions – is in line with the idea that the fronto-temporal language network stores our linguistic knowledge representations acquired through our experience with language (e.g. Fedorenko, 2014; Fedorenko, Nieto-Castañón, et al., 2012). These representations may take the form of sounds, sound combinations, morphemes, words, part-structures (Bybee, 1998; Jackendoff, 2002) or perhaps constructions (Goldberg, 1995). However, it certainly does not seem plausible that we would store representations that span multiple clauses. Thus, although the response to linguistic stimuli appears to increase as the stimuli match more and more closely the statistics of our prior linguistic experience, from a relatively low response to pseudoword sequences, to a stronger response to real words, to yet stronger responses to phrases and sentences (e.g. Fedorenko et al., 2010; Pallier et al., 2011; Snijders et al., 2009), the response appears to asymptote at the level of clauses/sentences, with larger, discourse-level, meanings/structures eliciting as strong a response as unconnected sentences.

Fourth, the fact that the domain-general MD brain regions responded more strongly during the *Non-Coherent* than the *Coherent* condition argues against the role of executive resources in discourse-level comprehension. A priori, the latter possibility seems plausible given that the MD system has been implicated in structuring diverse complex behaviours (e.g. Camilleri et al., 2018; Duncan, 2013; Müller, Langner, Cieslik, Rottschy, & Eickhoff, 2015). However, it appears that coherent discourses are not built by these domain-general brain regions.

And finally, the results reported here once again underscore the importance of paying attention to fine-grain functional distinctions among nearby brain regions rather than operating at the level of large anatomical areas like gyri and sulci. In particular, we observed a greater response to the *Coherent* condition in the two medial frontal ToM fROIs. However, nearby frontal MD fROIs showed the opposite pattern of response (see also Saxe, Schulz, & Jiang, 2006). Thus generalisations at the level of "MPFC" are unlikely to be correct.

More generally, a wide range of cognitive processes appear to co-localise to MPFC: from valuation and decision making, to memory, to multiple aspects of social cognition, etc. (Euston, Gruber, & McNaughton, 2012; Northoff et al., 2006; Roy, Shohamy, & Wager, 2012; Wagner, Haxby, & Heatherton, 2012, to name a few). One could argue that even in the presence of

multiple functionally distinct nearby regions, it may be productive to think about overarching organising principles / similar computations of the broader area (or the entire brain, for that matter). However, we believe that such attempts should not come at the cost of blurring potentially critical functional distinctions that may exist at a finer grain of analysis. Before any sweeping generalisations are made, it is critical to carefully and rigorously map out the detailed functional landscape using the kinds of individual-subject analyses we employ in our study, in order to understand which of the many functions that have been linked to the MPFC indeed arise in the "same" region, and which may originate in nearby distinct regions. Our study is one step in this direction: we focus on a ToM-sensitive region within the MPFC and provide evidence against its domain-specificity for ToM by showing that it responds to coherence in expository texts.

It is also important to note that although in some cases (including our study), the results of careful individual-subject-level analyses may be broadly consistent with those of related studies that have used traditional whole-brain group analyses (e.g. Ferstl & von Cramon, 2002; Lin et al., 2018) or meta-analytic approaches (e.g. Northoff et al., 2006), plenty of cases exist where group analyses have led to fundamentally flawed conclusions (e.g. Aguirre & Farah, 1998). Thus building on evidence that has emerged from group analyses can be risky (see Nieto-Castañón & Fedorenko, 2012), for a general discussion of sensitivity and functional resolution in traditional vs. fROI-based analyses).

In summary, the processing of discourse-level structure takes place within the Theory of Mind network: in particular, within its medial frontal component. Posterior ToM regions appear to only get engaged during text processing when the narratives require consideration of others' mental states. Thus, those posterior ToM regions do not support coherence processing in general or integration of information over longer time-windows (cf. Lerner et al., 2011). In contrast, the computations carried out by the medial frontal ToM regions are not limited to mental-content-rich narratives. However, the precise nature of the computations they perform within coherence building / discourse-level comprehension remains to be discovered.

Resources

All the study materials and methods can be found online in the following places:

(1) Critical task: materials; Statistical analysis inputs and code; random-effects *t*-map for the coherent > non-coherent contrast – https://osf.io/d86nc/

(2) ToM localiser – http://saxelab.mit.edu/superloc.php
(3) ToM parcels – http://saxelab.mit.edu/ToMgroupMaps.php
(4) Language and MD localisers – https://evlab.mit.edu/funcloc/download-paradigms
(5) Language and MD parcels – https://evlab.mit.edu/funcloc/download-parcels

ORCID

Nir Jacoby http://orcid.org/0000-0001-7936-9991
Evelina Fedorenko http://orcid.org/0000-0003-3823-514X

References

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*(1), 693–716.

Aguirre, G. K., & Farah, M. J. (1998). Human visual object recognition: What have we learned from neuroimaging? *Psychobiology*, *26*(4), 322–332.

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Barbey, A. K., Colom, R., & Grafman, J. (2013). Neural mechanisms of discourse comprehension: A human lesion study. *Brain*, *137*(1), 277–287.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, 6(5), 448–450.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain and Language*, 44(1), 80–120.

Beeman, M. J., Bowden, E. M., & Gernsbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. *Brain and Language*, 71(2), 310–336.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *The Journal of Neuroscience*, 17(1), 353–362.

Blank, I. A. (2016). *The functional architecture of language comprehension mechanisms: Fundamental principles revealed with fMRI* (N. Kanwisher & E. Fedorenko, Eds.). Massachusetts Institute of Technology.

Blank, I. A., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, 112(5), 1105–1118.

Blank, I., & Fedorenko, E. (in press). A new functional signature of high-level language regions: common timescales for integrating information.

Bybee, J. (1998). A functionalist approach to grammar and its evolution. *Evolution of Communication*, 2(2), 249–278.

Camilleri, J. A., Müller, V. I., Fox, P., Laird, A. R., Hoffstaedter, F., Kalenscher, T., & Eickhoff, S. B. (2018). Definition and characterization of an extended multiple-demand network. *Neuroimage*, 165, 138–147.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, 25(11), 4596–4609.

de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *The Journal of Neuroscience*, 36(24), 6553–6562.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a Theory of Mind task. *Neuroimage*, 55(2), 705–712.

Dufour, N., Redcay, E., Young, L., Mavros, P. L., Moran, J. M., Triantafyllou, C., et al. (2013). Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLoS ONE*, 8(9), e75468.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4), 172–179.

Duncan, J. (2013). The structure of cognition: Attentional episodes in mind and brain. *Neuron*, 80(1), 35–50.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23(10), 475–483.

Egidi, G., & Caramazza, A. (2016). Integration processes compared: Cortical differences for consistency evaluation and passive comprehension in local and global coherence. *Journal of Cognitive Neuroscience*, 28(10), 1568–1583.

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, 76(6), 1057–1070.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, 5, 335.

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39), 16428–16433.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions Lie side by side within Broca's area. *Current Biology*, 22(21), 2059–2062.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41), 16616–16621.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitefield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, 104(2), 1177–1194.

Fedorenko, E., Nieto-Castañón, A., & Kanwisher, N. (2012). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, 50(4), 499–513.

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, 18(3), 120–126.

Ferstl, E. C., Neumann, J., Bogler, C., & von-Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, 29(5), 581–593.

Ferstl, E. C., & von-Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research*, 11(3), 325–340.

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the fronto-median cortex contribute to language processing: Coherence or Theory of Mind? *Neuroimage*, 17(3), 1599–1612.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiack, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "Theory of Mind" in story comprehension. *Cognition*, 57(2), 109–128.

Friese, U., Rutschmann, R., Raabe, M., & Schmalhofer, F. (2008). Neural indicators of inference processes in text comprehension: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20(11), 2110–2124.

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 non-verbal tasks. *Neuropsychologia*, 38(1), 11–21.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, 5(8), 509–524.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48(1), 163–189.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175–204.

Haberlandt, K. (1980). Story grammar and reading time of story constituents. *Poetics*, 9(1-3), 99–118.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114(3), 357–374.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, 28(10), 2539–2550.

Hobbs, J. R. (1985). *On the coherence and structure of discourse.* Stanford, CA: Center for the Study of Language and Information.

Jackendoff, R. (2002). *Foundations of language.* Oxford: OUP.

Jacoby, N., Bruneau, E. G., Koster-Hale, J., & Saxe, R. (2016). Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *NeuroImage*, 126(C), 39–48.

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4), 2357–2364.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective & Behavioral Neuroscience*, 7(3), 153–191.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 115–126.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge: Cambridge University Press.

Kintsch, W., Mandel, T. S., & Kozminsky, E. (1977). Summarizing scrambled stories. *Memory & Cognition*, 5(5), 547–552.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *NeuroImage*, 33(1), 343–361.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8), 2906–2915.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.

Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., & Li, X. (2018). Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. *Cognitive, Affective & Behavioral Neuroscience*, 103(2), 1–11.

Maguire, E. A., Frith, C. D., & Morris, R. G. M. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain*, 122(Pt 10), 1839–1850.

Mahowald, K., & Fedorenko, E. (2016). Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *NeuroImage*, 139, 74–93.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, 62(1), 103–134.

Marcu, D. (2000). *The theory and practice of discourse parsing and summarization.* Cambridge, Massachusetts, United States: MIT Press.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1–71.

Mason, R. A., & Just, M. A. (2009). The role of the theory-of-mind cortical network in the comprehension of narratives. *Language and Linguistics Compass*, 3(1), 157–174.

Mason, R. A., & Just, M. A. (2011). Differentiable cortical networks for inferences concerning people's intentions versus physical causality. *Human Brain Mapping*, 32(2), 313–329.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In *The psychology of learning and motivation* (1st ed., Vol. 51, pp. 297–384). Elsevier.

Moss, J., & Schunn, C. D. (2015). Comprehension through explanation as the interaction of the brain's coherence and cognitive control networks. *Frontiers in Human Neuroscience*, 9, 107–117.

Moss, J., Schunn, C. D., Schneider, W., McNamara, D. S., & VanLehn, K. (2011). The neural correlates of strategic reading comprehension: Cognitive control and discourse comprehension. *NeuroImage*, 58(2), 675–686.

Müller, V. I., Langner, R., Cieslik, E. C., Rottschy, C., & Eickhoff, S. B. (2015). Interindividual differences in cognitive flexibility: Influence of gray matter volume, functional connectivity and trait impulsivity. *Brain Structure and Function*, 220(4), 2401–2414.

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, 63(3), 1646–1669.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14(9), 1105–1107.

Nijhof, A. D., & Willems, R. M. (2015). Simulating fiction: Individual differences in literature comprehension revealed with fMRI. *PLoS ONE*, 10(2), e0116492–17.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain – A meta-analysis of imaging studies on the self. *NeuroImage*, 31(1), 440–457.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), 2522–2527.

Palomero-Gallagher, N., Zilles, K., Schleicher, A., & Vogt, B. A. (2013). Cyto- and receptor architecture of area 32 in human and macaque brains. *The Journal of Comparative Neurology*, 521(14), 3272–3286.

Pinheiro, J. C., & Bates, D. M. (2000). Mixed-effects models in S and S-PLUS (Statistics and Computing). In J. Chambers, W. Eddy, W. Hardle, S. Sheather, & L. Tierney, (Eds.), New York: Springer-Verlag.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59–63.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, 72(5), 692–697.

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, 72(6), 675–688.

Rice, K., & Redcay, E. (2016). Interaction matters: A perceived social partner alters the neural processing of human speech. *NeuroImage*, 129, 480–488.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, 16(3), 147–156.

Sabbagh, M. A. (1999). Communicative intentions and language: Evidence from right-hemisphere damage and autism. *Brain and Language*, *70*(1), 29–69.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for One component of theory of mind. *Psychological Science*, *17*(8), 692–699.

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*(3-4), 284–298.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*, *4*(3), e4869.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34.

Singh, R., Fedorenko, E., Mahowald, K., & Gibson, E. (2016). Accommodating presuppositions Is inappropriate in implausible contexts. *Cognitive Science*, *40*(3), 607–634.

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *Journal of Neuroscience*, *34*(48), 15997–16008.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, *19*(7), 1493–1503.

Swett, K., Miller, A. C., Burns, S., Hoeft, F., Davis, N., Petrill, S. A., & Cutting, L. E. (2013). Comprehending expository texts: The dynamic neurobiological correlates of building a coherent text representation. *Frontiers in Human Neuroscience*, *7*, 853.

Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2015). Reading fiction and reading minds: the role of simulation in the default network. *Social cognitive and affective neuroscience*, *11*(2), 215–224.

Team, R. C. (2013). A language and environment for statistical computing.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic.

Vogt, B. A. (2016). Midcingulate cortex: Structure, connections, homologies, functions and diseases. *Journal of Chemical Neuroanatomy*, *74*, 28–46.

Wagner, D. D., Haxby, J. V., & Heatherton, T. F. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(4), 451–470.

Wochna, K. L., & Juhasz, B. J. (2013). Context length and reading novel words: An eye-movement investigation. *British Journal of Psychology (London, England : 1953)*, *104*(3), 347–363.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, *31*(2), 249–287.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage*, *25*(3), 1002–1015.

Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *NeuroImage*, *41*(4), 1408–1425.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia*, *48*(9), 2658–2664.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297.