Check for updates

Processing Speech and Thoughts during Silent Reading: Direct Reference Effects for Speech by Fictional Characters in Voice-Selective Auditory Cortex and a Theory-of-Mind Network

Ben Alderson-Day[1], Jamie Moffatt[1,2], Marco Bernini[1], Kaja Mitrenga[1]
Bo Yao[3], and Charles Fernyhough[1]

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Abstract

■ Stories transport readers into vivid imaginative worlds, but understanding how readers create such worlds—populating them with characters, objects, and events—presents serious challenges across disciplines. Auditory imagery is thought to play a prominent role in this process, especially when representing characters' voices. Previous research has shown that direct reference to speech in stories (e.g., *He said, "I'm over here"*) may prompt spontaneous activation of voice-selective auditory cortex more than indirect speech [Yao, B., Belin, P., & Scheepers, C. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience, 23*, 3146–3152, 2011]. However, it is unclear whether this effect reflects differential processing of speech or differences in linguistic content, source memory, or grammar. One way to test this is to compare direct reference effects for characters speaking and thinking in a story. Here, we present a multidisciplinary fMRI study of 21 readers' responses to characters' speech and thoughts during silent reading of short fictional stories. Activations relating to direct and indirect references were compared for both speaking and thinking. Eye-tracking and independent localizer tasks (auditory cortex and theory of mind [ToM]) established ROIs in which responses to stories could be tracked for individuals. Evidence of elevated auditory cortex responses to direct speech over indirect speech was observed, replicating previously reported effects; no reference effect was observed for thoughts. Moreover, a direct reference effect specific to speech was also evident in regions previously associated with inferring intentions from communication. Implications are discussed for the spontaneous representation of fictional characters and the potential roles of inner speech and ToM in this process. ■

INTRODUCTION

Stories can conjure complex imaginative worlds that offer immersion and transportation for the reader (Green, 2004; Green, Brock, & Kaufman, 2004; Ryan, 1999; Gerrig, 1993). Fictional characters in particular are sometimes experienced with a vividness and complexity, which can linger beyond the page (Alderson-Day, Bernini, & Fernyhough, 2017; Maslej, Oatley, & Mar, 2017). Understanding how these experiences are created by the mind—often with apparent automaticity and spontaneity—is a challenge for a wide range of disciplines beyond psychology, including literary theory, narratology, philosophy of mind, and cognitive neuroscience (Herman, 2013). Far from passively "receiving" information from the writer, readers actively and creatively engage with fictional texts in a way that draws on multiple psychological resources (Polvinen, 2016; Caracciolo, 2014; Kukkonen, 2014; Oatley, 2011; Bortolussi & Dixon, 2003).

One approach to understanding the qualitative richness of the reading experience has been to study inner speech (sometimes also referred to as inner monologue or articulatory imagery; Alderson-Day & Fernyhough, 2015; Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Lœvenbruck, 2014). Intuitively, reading is often associated with the sounding out of an "inner voice," and the self-reports of readers involve various kinds of auditory imagery when engaged in a story (Vilhauer, 2016). Although the reliability of readers' introspective reports has been questioned (Caracciolo & Hurlburt, 2016), empirical evidence of inner speech involvement during silent reading is well documented (Filik & Barber, 2011; Alexander & Nygaard, 2008). Moreover, silent reading appears to elicit activity in perisylvian regions and auditory association cortex (Magrassi, Aromataris, Cabrini, Annovazzi-Lodi, & Moro, 2015; Perrone-Bertolotti et al., 2012), particularly when characters' voices and speech are being described

[1]Durham University, [2]University of Sussex, [3]University of Manchester

© 2020 Massachusetts Institute of Technology. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Journal of Cognitive Neuroscience 32:9, pp. 1637–1653
https://doi.org/10.1162/jocn_a_01571

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

(Brück, Kreifelts, Gößling-Arnold, Wertheimer, & Wildgruber, 2014; Yao, Belin, & Scheepers, 2011). Such findings have been taken as evidence of the reading experience—and its evocation of inner speech—being almost akin to hearing external voices (Petkov & Belin, 2013).

A good example of this is provided by texts involving direct speech. When direct reference is made to a character overtly speaking in a text (*he said, "the cat is over there"*), it is thought to evoke a more vivid experience of the storyworld than if the same overt speech is only indirectly referred to (*he said that the cat is over there*). It has been suggested that the purpose of such constructions is to demonstrate (and thus depict) a situation, rather than merely describe it (Clark & Gerrig, 1990). Evidence that this could resemble hearing an actual voice is provided by Yao et al. (2011), who compared fMRI responses in auditory cortex for participants silently reading short stories that contained either direct or indirect reference to speech. Although both kinds of speech activated auditory cortex, direct speech was associated with a greater response than indirect speech in voice-selective regions of the right superior and middle temporal lobe (as defined by a separate auditory localizer task; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). As the stories were very short (three to four sentences in total) and participants were not prompted to imagine the voices, characters, or stories in any specific way, this suggests that fairly minimal textual markers for direct speech can elicit a response in cortical regions that are selective for voice perception.

If direct speech in text can prompt this kind of reaction, a second question is why readers appear to respond in this way. In a separate study, Yao, Belin, and Scheepers (2012) observed similarly enhanced responses in voice-selective regions for direct speech quotations when they were being read by a monotonous voice. Building on Barsalou's theory of embodied cognition (Barsalou, 2008), they suggested that auditory cortex activation may have a role in constructing a perceptual simulation of the emotional prosody and intonation of the speaker's voice, given that such information is either absent or diminished in the case of both silent reading and monotonous listening. This would not rule out perceptual simulation during other kinds of silent reading but characterizes direct reference as a cue to simulate suprasegmental and communicative properties of speech from text (Yao et al., 2011, 2012).

The effects of direct speech and its potential consequences for simulation can be questioned, however. If direct speech prompts more vivid imagery or provides more communicative information (e.g., tone or emotional content), this would plausibly be reflected in reader comprehension. However, in a series of behavioral experiments, Eerland, Engelen, and Zwaan (2013) reported inconsistent evidence for either perceptual or communicative information being more available to readers after direct speech quotations. Instead, they suggested that the use of direct quotations prompts better memory for the verbatim content of characters' utterances, whereas indirect speech assists the building of a situation model, that is, an overall "representation of the referential situation" (Eerland et al., 2013, p. 7; van Dijk & Kintsch, 1983). Supporting this, source memory for characters' utterances is actually enhanced for indirect, not direct, speech quotations (Eerland & Zwaan, 2018)—suggesting that the potential vividness of direct speech is not used for tracking information about who said what (or could even obstruct such tracking, when compared to indirect speech). Finally, the typographical and grammatical differences between direct and indirect speech make it difficult to clearly compare their specific consequences for mental simulation. Along with potentially alerting the reader to pay attention to text, direct sentences are typically shorter than indirect sentences, are syntactically simpler, and may be expected to prompt changes in reader perspective (Köder, Maier, & Hendriks, 2015; Coulmas, 2011; Clark & Gerrig, 1990). As such, the effect of direct speech on the reader, and its potential function in the imaginative response of reading, remains unclear.

One way to explore this topic—in a way that might begin to address some of the above concerns—is to compare references to characters' speech with another kind of representation that fictional narratives can involve: characters' thoughts. Although theories of mental simulation during reading emphasize various forms of sensory and embodied simulation (e.g., Kurby & Zacks, 2013; Zwaan, Madden, Yaxley, & Aveyard, 2004), fictional narratives have been proposed to place specific sociocognitive demands on the reader (Mar & Oatley, 2008; Zunshine, 2006). Typically, a reader must track the mental states of multiple characters, following their beliefs, intentions, and desires through a narrative, to make sense of actions, decisions, and responses to events in the storyworld (Spreng, Mar, & Kim, 2009; Herman, 2008; Palmer, 2004; Gerrig, Brennan, & Ohaeri, 2001),[1] all of which imply a central role for theory of mind (ToM) in the reading process.

How might this shed light on direct speech? First is because it provides a contrasting example of direct reference. Both indirect and direct references to thinking are used in narratives. Indirect thought, which is usually considered the representational norm (Leech & Short, 2007, p. 268), is more flexible and can be used to represent verbal, preverbal, and nonverbal mental processes from the perspective of the character (e.g., he thought that *X*; he felt that *Y*; he was willing to do *Z*). Direct thought (also referred to as "quoted monologue"; Cohn, 1978) is used to represent, verbatim, the linguistic silent articulation of verbal thoughts (He thought "this is so complicated!").

The verbalized nature of depicting characters' thoughts is almost identical in form and complexity to direct speech (i.e., when used in a basic form; indirect thoughts in more

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

extended narratives can be used in highly complex ways). Contrasting these forms of speech and verbal thought can therefore provide a test of Yao et al.'s (2011) interpretation of direct reference effects by assessing how specific they might be to vocal information, while at the same time controlling for typographic features. If Yao et al.'s conjecture is correct, direct reference to speech—but not necessarily thoughts—may be expected to elicit recruitment of voice-selective regions of auditory cortex, to specifically simulate the perceptual qualities of characters speaking out loud in the storyworld. In contrast, if direct speech and direct thought elicit similar responses, then a voice-specific account of direct reference would be harder to maintain. It could be the case that both speech and thoughts elicit some form of perceptual simulation under direct reference, but they would be doing so despite clear dissimilarities in the auditory scenario (one is an external utterance, and the other is a form of internal monologue). Instead, showing that direct reference to speech and thoughts prompts a generally greater response in auditory cortex could support alternative interpretations: It may, for example, simply reflect a greater level of engagement that happens when quote marks prompt the reader to pay attention to verbatim content (Eerland et al., 2013). In that scenario, there would be nothing special about speech for understanding direct reference effects.

Second, exploring sociocognitive processing and contrasting how this works for speaking and thinking are potentially highly informative for understanding direct speech effects. ToM has multiple components, each with its own developmental trajectory (Fernyhough, 2008; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Whereas some socio-cognitive skills are evident early in infancy—such as the ability to follow others' attentional cues (Behne, Liszkowski, Carpenter, & Tomasello, 2012; Woodward, 1998)—the ability to cognitively represent others' mental states when incorrect is thought to emerge later in childhood (typically around 4 years old; Wellman, Cross, & Watson, 2001). Similarly, understanding pragmatic information and speaker intention from prosody shows a competence–performance gap, in which vocal cues to emotion are recognized very early in infancy but are only used consistently (in the face of, e.g., conflicting cues) in older children (Esteve-Gibert & Guellaï, 2018). ToM and social cognition more generally are associated with a canonical network of regions in the medial pFC (mPFC), precuneus, and TPJ bilaterally (Molenberghs, Johnson, Henry, & Mattingley, 2016; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Saxe & Kanwisher, 2003; Fletcher et al., 1995). Of these, representing the thoughts and intentions of others in particular has been argued to localize to regions of the TPJ and precuneus (Schurz, Tholen, Perner, Mars, & Sallet, 2017; Saxe & Powell, 2006), whereas mPFC has been linked to the processing of more constant traits associated with self and other (van Overwalle, 2009).

If direct speech prompts a detailed simulation of suprasegmental vocal information (such as emotional tone or prosody), then this may also be reflected in social–cognitive regions—specifically for areas associated with interpreting or reasoning about a speaker's communicative intentions. For example, using a nonverbal, cartoon-based story task, Ciaramidaro et al. (2007) observed that bilateral TPJ regions in particular are associated with tracking different kinds of intent associated with socio-communicative interactions. If direct speech prompted similar activation, this would support an extension of Yao et al.'s (2011) original theory to suggest that direct reference involves constructing a broader, sociopercep-tual simulation than merely how a voice sounds. Contrastingly, if tracking characters and their intentions is an ultimately separate process from simulating the perceptual features of characters' voices, then no direct effect for speech would necessarily be expected in ToM regions. Instead, it is possible that references to characters' mental states—but not their speech—would be most likely to engage such regions, irrespective of any direct reference effect.

To investigate this, we adapted Yao et al.'s (2011) paradigm to include direct and indirect references to characters' verbal thoughts and speech in a 2 × 2 design. We used eye tracking and an auditory localizer task to study cortical responses specific to each individual's reading times and voice-selective regions. To explore the broader effect of direct speech in regions commonly associated with inferring communicative intentions, we also included a version of Ciaramidaro et al.'s (2007) story task as a second localizer. Many standard ToM tasks use written short stories in which characters' false beliefs must be inferred from textual information, but using such stories could be expected to overlap considerably with other reading tasks (in terms of both stimuli and task demands). Instead, by using a wordless, cartoon-based ToM task, we could avoid this potential confound with the demands of our main direct/indirect story task. On the basis of the original findings of Yao et al. (2011, 2012), we hypothesized that (i) direct reference effects would be evident for speech but not thoughts in auditory cortex. In accordance with the claim that this facilitates prosodic and communicative processing of the utterance, we also predicted that (ii) the voice-specific effect of direct reference would extend to ToM-related regions. In contrast, no direct reference effects were expected for thoughts, in either network.

METHODS

Participants

An initial sample of 30 individuals took part in the full MRI procedure, but nine participants did not produce a full data set because of the following exclusions (one incidental finding, one insufficient accuracy [<60%] on the

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

story task, two no clear voice-selective response on the auditory localizer task, five insufficient eye-tracking data; three male, six female). As such, analysis proceeded with a final sample of 21 (age: M = 23.49, SD = 6.63; three male, 18 female). All participants were right-handed, native English speakers, with normal or corrected-to-normal vision. All procedures were approved by a university ethics subcommittee.

Measures

Story Task

Following Yao et al. (2011), participants viewed a series of short stories containing two preparation sentences (Sentences 1 and 2) and a target sentence, containing a character either (i) speaking or thinking with (ii) direct or indirect reference. On each trial, participants viewed a fixation cross for 1–2 sec (jittered at random), followed by one slide per sentence, presented sequentially (see Figure 1). Viewing times per slide were determined using the following formula: (Words × 100 msec) + (Syllables × 50 msec) + 2000 msec. Mean presentation times were 5.61 and 5.72 sec for Sentences 1 and 2, respectively, and 5.95 and 6.22 sec for direct and indirect target sentences, respectively, reflecting the slightly longer length of indirect sentences on average (18.6 words per indirect sentence compared to 16.8 for direct sentences). To allow for sufficient trials in each condition, the number of stories was increased from the 90 trials used in Yao et al. (2011) to 120, split across two 20-min runs (additional stories were prepared by a narratologist, M. B., to follow the length, complexity, and style of the original stimuli and ensure balance across the four conditions). Each run also contained three 30-sec break periods, occurring every 20 trials. An attentional check (a simple comprehension question relating to factual content from the preceding story) was included after 25% of trials, with participants having 6 sec to respond.[2] Four random orders of trials were generated, counterbalancing the combination of voice/thought and direct/indirect target sentences across participants. Eye-tracking timings were collected as an indicator of participants' reading responses for the two preparation sentences and the target sentence. Specifically, participants' first

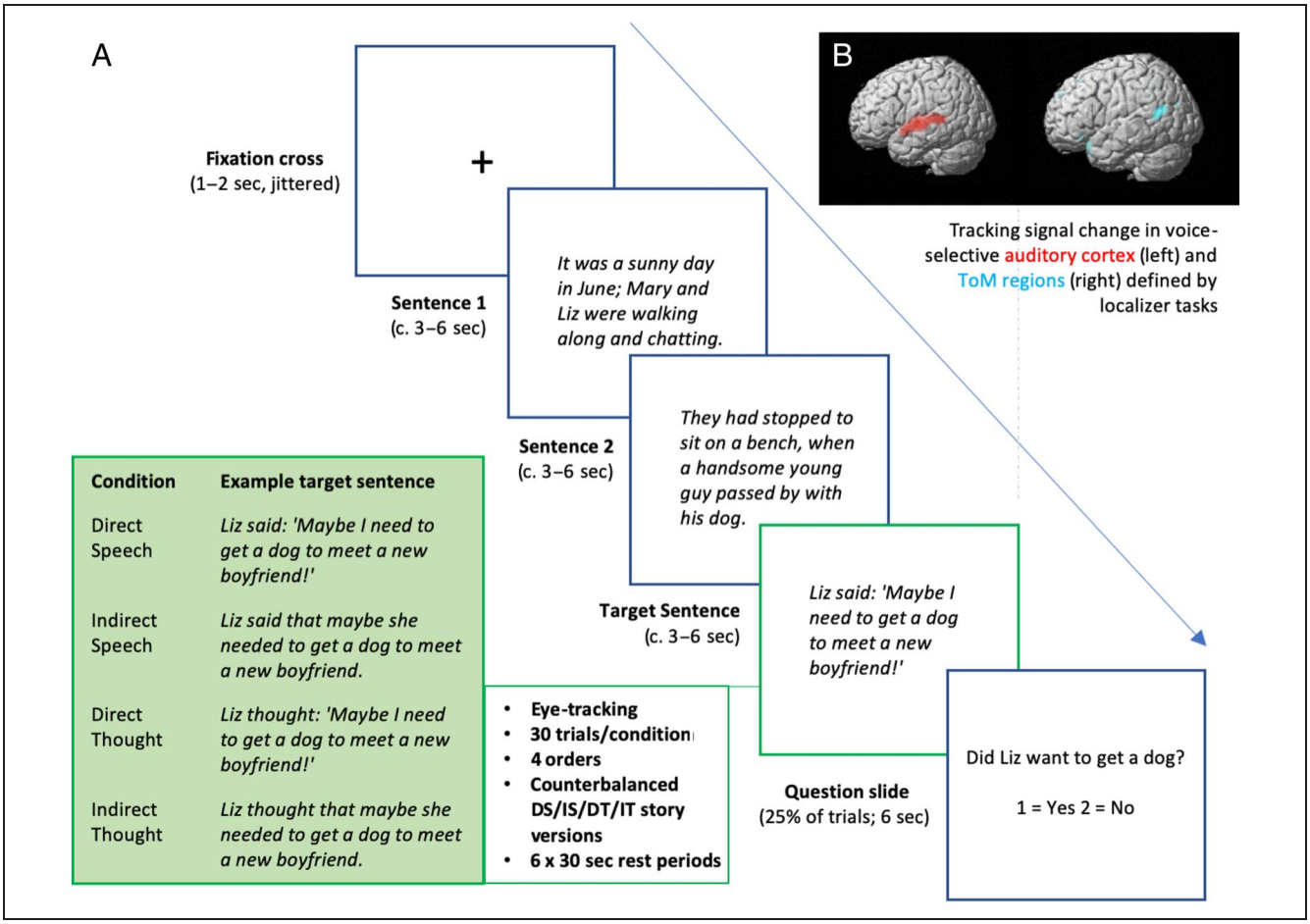


Figure 1. Adapted story task (A) with direct reference to voices and thoughts applied to auditory and ToM localizer regions (B). Figure 1B depicts left-sided sagittal view (rendered, $p < .05$, FWE); note that auditory and ToM regions were observed bilaterally (see Table 1).

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Table 1. Accuracy Rates, Reading Onsets, and Reading Times by Task Condition

	Direct Speech		*Indirect Speech*		*Direct Thought*		*Indirect Thought*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Accuracy (%)	79.84	18.43	79.21	16.08	85.56	15.87	86.35	15.90
Reading onset (sec)	0.56	0.32	0.57	0.33	0.61	0.36	0.67	0.36
Duration (sec)	4.02	0.54	4.10	0.54	4.06	0.44	4.11	0.58

fixation (the beginning of the sentence) and last fixation (the final line of the target sentence) within the text area were used to define reading onsets and offsets of characters' speaking and thinking in the target sentence. These were then directly included in the fMRI model to account for individual differences in the reading response.

Auditory Localizer Task

The auditory localizer task was identical to that used in Yao et al. (2011). Participants listened to 20 blocks of vocal stimuli and 20 blocks of nonvocal stimuli, along with 20 silent blocks that were used as a baseline. The blocks were presented randomly. Each block was 8 sec long, and the task lasted 10 min. The contrasting brain activity in response to the vocal and nonvocal stimuli reliably localizes voice-selective areas of the auditory cortex (Yao et al., 2011; Belin et al., 2000).

ToM Task

The cartoon-based ToM task was adapted from a task used by Walter et al. (2004) and Ciaramidaro et al. (2007). Participants viewed a sequence of three cartoon story vignettes ("story" phase) and were required to indicate a logical end of each story based on the three presented images ("choice" phase). The story phase included either reasoning about characters' intentions when communicating with others (e.g., a man indicating whether a seat is free on a train) or physical reasoning (e.g., a water pipe bursting). The images were displayed sequentially for 3 sec in the story phase and for 7 sec in the choice phase. The intertrial intervals lasted between 7 and 11 sec. In total, 10 ToM stories and 10 physical reasoning stories were presented in a random order. Participants answered (A, B, or C) by a button press. The task took 9 min to complete. The contrasting brain activity in response to the ToM reasoning stories compared to physical reasoning stories has been observed to prompt activity in brain regions often associated with ToM, including the right TPJ, precuneus, and anterior paracingulate cortex (Alderson-Day et al., 2016; Ciaramidaro et al., 2007; Walter et al., 2004).

Data Acquisition

fMRI data were acquired at Durham University Neuroimaging Centre using a 3-T Magnetom Trio MRI system (Siemens Medical Systems) with standard gradients and a 32-channel head coil. T2*-weighted axial EPI scans were acquired with the following parameters: field of view = 212 mm, flip angle = 90°, repetition time = 2000 msec, echo time = 30 msec, number of slices = 32, slice thickness = 3.0 mm, interslice gap = 0.3 mm, and matrix size = 64 × 64. Story task data were collected across 2 × 20-min runs consisting of 600 volumes each; auditory and ToM tasks took roughly 10 min each and consisted of 300 and 281 volumes, respectively. The first three volumes of each EPI run were discarded to allow for equilibrium of the T2 response. For each participant, an anatomical scan was acquired using a high-resolution T1-weighted 3-D sequence (number of slices = 192, slice thickness = 1 mm, matrix size = 512 × 512, field of view = 256 mm, echo time = 2.52 msec, repetition time = 2250 msec, flip angle = 9°). Eye-tracking data were collected using a LiveTrack system (Cambridge Research Systems) with MATLAB 2016b (The Mathworks, Inc.).

Data Analysis

All MRI analyses were conducted using SPM Version 12 (Wellcome Department of Cognitive Neurology) implemented in MATLAB. Images were slice-time corrected before being realigned to the first image to correct for head movement. Volumes were then normalized into standard stereotaxic anatomical Montreal Neurological Institute space using the transformation matrix calculated from the first EPI scan of each participant and the EPI template. The default settings for normalization in SPM12 and the standard EPI template supplied with SPM12 were used. The normalized data with a resliced voxel size of 2 × 2 × 2 mm were smoothed with an 8-mm FHWM isotropic Gaussian kernel to accommodate intersubject anatomical variation. The time-series data were high-pass filtered with a high-pass cutoff of 1/128 Hz, and first-order autocorrelations of the data were estimated and corrected for. Movement parameters from the realignment phase were visually inspected for outliers and included as regressors for single-participant (first-level)

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Table 2. Whole-Brain Coordinates for (A) Auditory and (B) ToM Localizer Tasks

Location	*x*	*y*	*z*	*k*	*t*	*z*	p_{FWE}
(A) Auditory (vocal > nonvocal)							
L middle temporal gyrus	−60	−14	−2	3512	17.08	7.34	<.001
L superior temporal gyrus	−58	−2	−4		11.67	6.34	
L middle temporal gyrus	−60	−36	6		11.42	6.29	
R superior temporal gyrus	56	−18	0	5051	12.76	6.58	<.001
R temporal pole	48	12	−18		11.85	6.39	
R superior temporal gyrus	64	−4	0		10.58	6.08	
(B) ToM (communicative inference > physical reasoning)							
R middle cingulate cortex	4	−56	40	1674	12.43	5.95	<.001
L precuneus	−2	−50	46		10.25	5.10	
WM	−14	−50	34		9.51	5.33	
R middle temporal gyrus	48	−50	24	5835	12.24	5.92	<.001
R superior temporal gyrus	58	−46	20		12.03	5.88	
R middle temporal gyrus	56	2	−22		10.44	5.55	
WM	−36	−56	16	5472	11.59	5.79	<.001
L temporal pole	−38	22	−22		11.45	5.76	
L middle temporal gyrus	−62	−56	18		10.35	5.53	
L superior medial frontal gyrus	−8	32	52	4021	10.68	5.60	<.001
R superior medial frontal gyrus	4	54	36		9.21	5.26	
L posterior-medial frontal gyrus	−4	14	60		7.92	4.90	
L gyrus rectus	0	52	−14	279	7.76	4.85	.015

All results $p < .05$, FWE, at peak/cluster level. Minimum cluster size: $k = 50$. L = left; R = right; ToM = theory of mind; WM = white matter.

analyses. ROI analyses were conducted using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002). Individual ROIs were defined using $p < .05$ corrected for FWE at cluster level, in temporal cortical regions for the auditory localizer task and clusters in mPFC, precuneus, and TPJ regions for the ToM task. Where significant clusters were not evident for individual participants at this level, a more liberal threshold of $p < .001$ (uncorrected) was used to maximize sensitivity to individual differences; participants who showed no clusters in these regions even at the more liberal threshold were excluded from analyses (two auditory, six ToM). All whole-brain analyses are presented at $p < .05$, FWE, cluster-level corrected. All statistical analyses of mean beta values were conducted using R and jamovi; figures were generated using ggplot2 and MicroGL. Effect sizes are reported as Cohen's d for pairwise comparisons and η_p^2 values for ANOVA main and interaction effects. η_p^2 values can be considered as small, moderate, and large effects with values of .099, .0588, and .1379, respectively (Richardson, 2011; Cohen, 1969).

RESULTS

Accuracy on the task was generally high ($M = 82.5\%$, $SD = 7.4\%$) indicating that participants maintained attention despite the 40-min duration of the task. Repeated-measures ANOVA with a 2 × 2 (Form × Reference) design was used to compare behavioral responses for the four conditions (see Table 1). No main effects, interaction effects, or pairwise comparisons were significant for condition accuracy, although we observed a nonsignificant trend for participants to be slightly less accurate on speech trials compared with thought trials, $F(1, 20) = 3.34$, $p = .082$, $\eta_p^2 = .14$ ($p > .14$ for all other effects and comparisons). For the duration of reading times, the only effect close to significance was for direct compared with indirect reference, $F(1, 20) = 3.57$, $p = .073$,

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Table 3. Whole-Brain Coordinates for Speech and Thought Sentences vs. Baseline

Location	*x*	*y*	*z*	*k*	*t*	*z*	p_{FWE}
Direct speech							
L middle temporal gyrus	−50	−26	−6	1198	11.25	6.25	<.001
L middle temporal gyrus	−62	−18	−8		9.66	5.83	
L middle temporal gyrus	−56	−46	4		8.55	5.49	
R temporal pole	48	14	−24	409	10.92	6.17	<.001
R temporal pole	44	22	−28		9.21	5.69	
L temporal pole	−48	20	−14	895	8.83	5.58	<.001
L inferior frontal gyrus	−50	22	20		8.64	5.51	
L inferior frontal gyrus	−46	16	24		8.31	5.41	
L superior medial frontal gyrus	−10	54	28	124	8.63	5.51	<.001
L precentral gyrus	−44	−2	54	112	8.43	5.45	<.001
Indirect speech							
L middle temporal gyrus	−54	−34	−2	550	11.84	6.38	<.001
L middle temporal gyrus	−50	−26	−4		10.42	6.04	
L middle temporal gyrus	−60	−20	−6		7.73	5.2	
L temporal pole	−50	12	−22	292	9.35	5.74	<.001
L inferior frontal gyrus	−48	28	−8		7.57	5.14	
L temporal pole	−46	22	−14		7.52	5.12	
R medial temporal pole	50	16	−28	308	9.31	5.73	<.001
R temporal pole	44	22	−26		8.46	5.45	
L middle temporal gyrus	−52	−52	18	144	7.87	5.25	<.001
L inferior frontal gyrus	−54	20	22	76	7.79	5.22	<.001
L middle temporal gyrus	−54	−34	−2	550	11.84	6.38	<.001
L middle temporal gyrus	−50	−26	−4		10.42	6.04	
L middle temporal gyrus	−60	−20	−6		7.73	5.2	
Direct thought							
L middle temporal gyrus	−54	−34	0	689	10.79	6.13	<.001
L middle temporal gyrus	−50	−26	−4		8.6	5.5	
L middle temporal gyrus	−62	−18	−8		8.31	5.41	
R medial temporal pole	50	10	−24	344	9.6	5.81	<.001
L superior medial frontal gyrus	−8	56	28	65	8.75	5.55	<.001
L superior frontal gyrus	−10	58	38		6.69	4.79	
R middle temporal gyrus	52	−36	−2	50	8.16	5.35	<.001
L inferior frontal gyrus	−54	20	24	53	7.18	4.99	<.001

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Table 3. (*continued*)

Location	*x*	*y*	*z*	*k*	*t*	*z*	p_{FWE}
Indirect thought							
L middle temporal gyrus	−54	−34	−2	1048	11.2	6.23	<.001
L middle temporal gyrus	−50	−26	−6		11.07	6.2	
L middle temporal gyrus	−54	−54	16		9.74	5.85	
R medial temporal pole	50	12	−24	343	10.78	6.13	<.001
R temporal pole	46	20	−26		10.4	6.03	
R middle temporal gyrus	50	−38	−2	56	8.91	5.6	<.001
L superior medial gyrus	−10	54	28	76	8.89	5.59	<.001
L precentral gyrus	−42	−2	56	70	8.52	5.48	<.001
L temporal pole	−52	12	−20	206	8.42	5.44	<.001
L temporal pole	−44	18	−16		8.42	5.44	
L temporal pole	−46	16	−32		6.69	4.79	

All results $p < .05$, FWE, at cluster and peak levels. Minimum cluster size: $k = 50$. L = left; R = right; ToM = theory of mind; WM = white matter.

$\eta_p^2 = .15$, which likely reflected the slightly longer lengths of indirect sentences. All other effects and comparisons for duration were also nonsignificant (all *p*s > .15). Reading onsets, in contrast, showed a main effect of Form, $F(1, 20) = 7.10, p = .015, \eta_p^2 = .26$, such that readers were quicker to start reading speech trials; follow-up pairwise comparisons indicated that this was only significantly quicker for direct speech compared with indirect thought, $t = 2.20$, $df = 36.24$, $p = .035$ (uncorrected), $d = 0.4$, all other *p*s > .10.

Whole-brain analyses—included here for descriptive purposes—indicated that the vocal > nonvocal contrast from the auditory localizer task was associated with significantly greater activation in bilateral auditory cortices, across the middle temporal gyrus (MTG) and superior temporal gyrus (see Figure 1B and Table 2). Compared with baseline, each of the four reading task conditions was associated with temporal activation bilaterally, with the largest clusters being observed along the dorsal bank of the left MTG (Table 3).

Responses to Characters' Speech and Thoughts in Voice-Selective Auditory Cortex

A repeated-measures ANOVA was used to compare mean beta values in auditory ROIs for story passages containing characters' speech or thoughts (i.e., Form) in direct or indirect reference, in a 2 × 2 design. No significant main effect of Form was evident, $F(1, 20) = 0.31$, $p = .584$, $\eta_p^2 = .02$, although a trend was observed for reference in favor of direct quotation, $F(1, 20) = 4.00$, $p = .059$, $\eta_p^2 = .17$. The interaction of Form and Reference was significant, $F(1, 20) = 7.08, p = .015, \eta_p^2 = .26$. As displayed in Figure 2, this was largely driven by a specific direct reference effect for character's speech, but not thoughts. Pairwise comparisons indicated that mean beta values for direct speech were significantly higher than those for indirect speech ($p = .006$, $d = 0.84$, Bonferroni corrected), but no other pairwise contrasts were significant (all *p*s > .25).

Responses to Speech and Thoughts in a ToM Network

We then applied the same analyses to responses in a ToM network identified via the cartoons task. As shown in

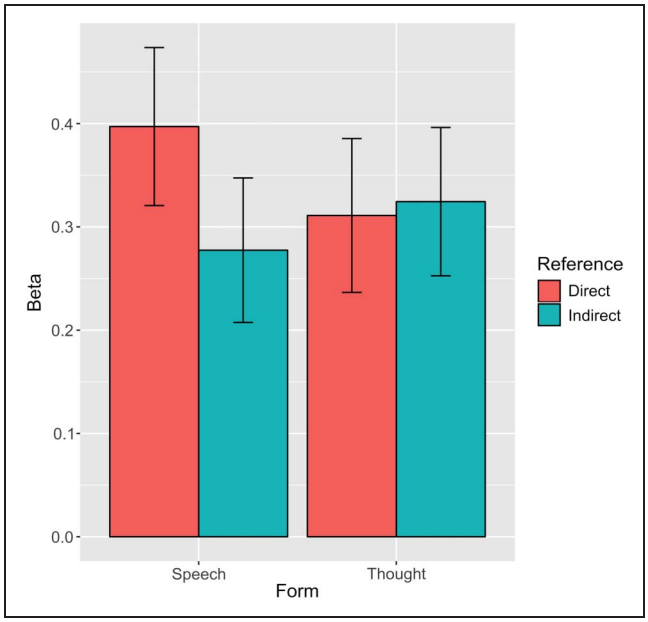


Figure 2. Mean beta values for direct and indirect references to characters' speech and thoughts in voice-selective auditory cortex. Error bars represent *SEM*.

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

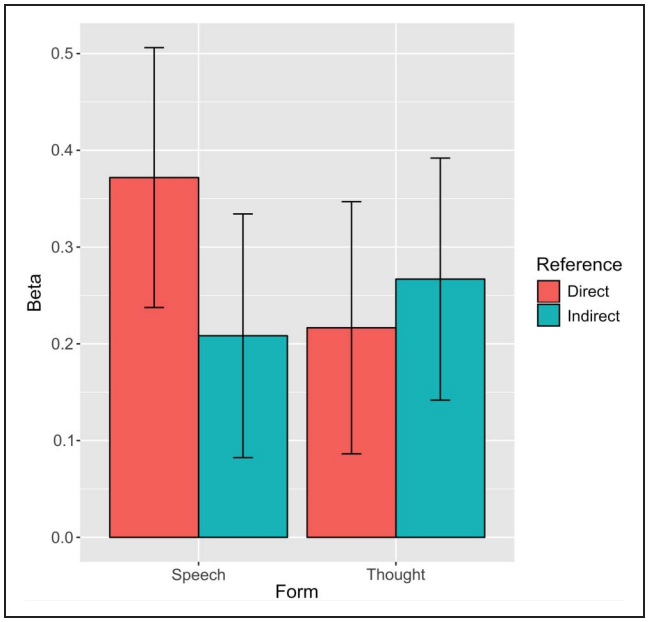


Figure 3. Mean beta values for direct and indirect references to characters' speech and thoughts in ToM regions. Error bars represent *SEM*.

Table 1, a range of typical regions were identified in the contrast between communicative inference reasoning and physical reasoning on the task, including the mPFC, precuneus, and TPJ bilaterally. Sixteen of the 21 individuals produced ToM networks with significant clusters in at least one of these regions, and their beta values were taken forward for ROI analysis (15/16, right TPJ; 12/16, left TPJ; 7/16, precuneus; 6/16, mPFC). When the mean beta values were compared in these areas in a repeated-measures ANOVA, no main effects of Form, $F(1, 15) = 0.49$, $p = .493$, $\eta_p^2 = .03$, or Reference, $F(1, 15) = 1.74$, $p = .207$, $\eta_p^2 = .10$, were observed, but a significant interaction was again evident, $F(1, 15) = 9.39$, $p = .008$, $\eta_p^2 = .38$.[3] As Figure 3 shows, this too was driven by responses for direct speech (compared with indirect speech), and this was the only significant difference between the conditions ($p = .016$, $d = 0.90$, Bonferroni corrected).

We then conducted an exploratory whole-brain analysis to investigate any further potential differences for direct versus indirect speech. Significant increases in signal for direct over indirect speech were evident in three regions: right TPJ (encompassing right angular gyrus [AG] and MTG), left inferior frontal gyrus (IFG), and left superior parietal lobule (see Figure 4). Using the online meta-analytic tool Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), the most common functional terms associated with these regions were "network DMN" for the right AG (posterior probability = 0.73), "theory mind" for the right MTG ($p = .88$), "semantic" for the left IFG ($p = .88$), and "imagery" for the left SPL ($p = .78$). Despite the apparent direct speech effect in voice-selective regions of the auditory cortex, no significant increase in signal was seen for this region when correcting across the whole brain for direct versus indirect speech (see Table 3). No regions were more active in the reverse contrast (indirect > direct speech).

Other exploratory whole-brain comparisons indicated few differences between conditions. Two exceptions were direct speech versus direct thought and direct reference versus indirect reference (i.e., with speech and thought sentences combined). Direct speech compared to direct thought was associated with greater activation in the right insula and anterior and middle cingulate, including regions bordering on the pre-SMA (see Table 4B). Direct reference was observed to predominantly activate occipital and parietal regions more than indirect reference (Table 4C). Their reverse contrasts (direct thought > direct speech; indirect > direct) produced no significant clusters, even at an uncorrected

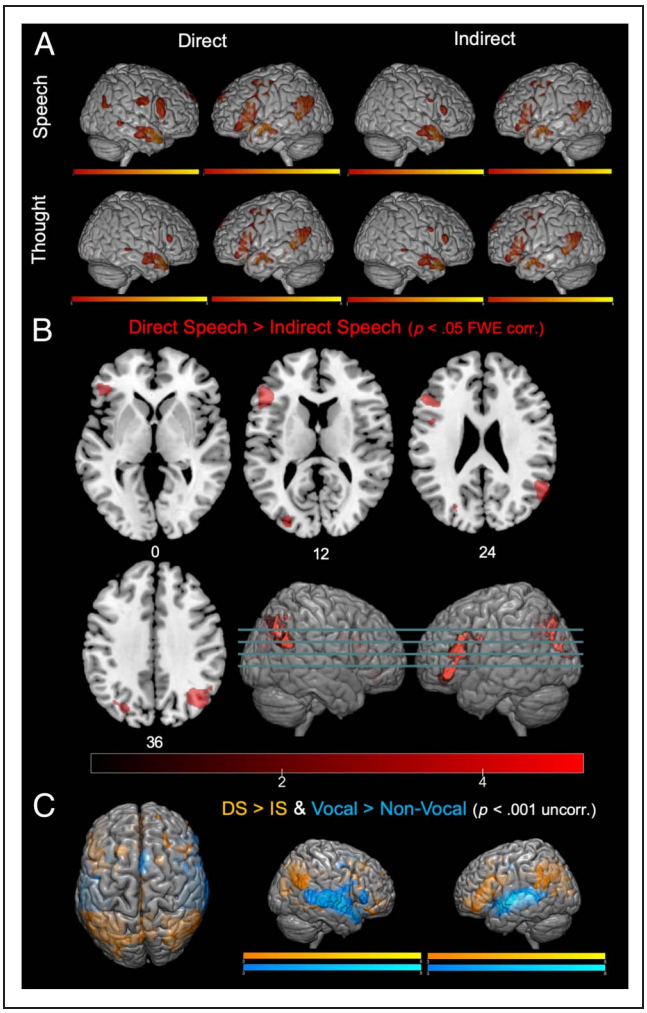


Figure 4. Whole-brain activation for (A) direct and indirect speech and thoughts versus baseline, (B) direct > indirect speech (DS > IS), and (C) DS > IS with group-level vocal > nonvocal contrast from auditory localizer data. All activations significant at cluster level, $p < .05$, FWE, apart from C, which reflects $p < .001$, uncorrected (uncorr.). corr. = corrected.

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Table 4. Whole-Brain Activation Differences for (A) Direct vs. Indirect Speech, (B) Direct Speech vs. Direct Thought, and (C) Direct vs. Indirect Reference

Location	*x*	*y*	*z*	*k*	*t*	*z*	p_{FWE}
(A) Direct speech > indirect speech							
R angular gyrus	42	−66	36	920	5.64	4.31	<.001
R angular gyrus	40	−72	42		5.54	4.26	
R middle temporal gyrus	56	−56	24		5.30	4.14	
L inferior frontal gyrus	−50	30	14	865	5.30	4.14	<.001
L inferior frontal gyrus	−52	26	22		5.23	4.11	
L inferior frontal gyrus	−44	32	0		4.66	3.79	
L superior parietal lobule	−24	−78	48	413	4.69	3.81	.007
L middle occipital gyrus	−28	−90	12		4.67	3.80	
L middle occipital gyrus	−26	−80	18		4.19	3.51	
(B) Direct speech > direct thought							
R middle cingulate cortex	14	26	32	438	5.36	4.17	.002
R ACC	6	36	28		4.92	3.93	
L ACC	−10	28	28		4.55	3.73	
Right insula	34	18	−6	231	5.26	4.12	.036
Right insula	28	24	−10		5.20	4.09	
(C) Direct > indirect							
L superior parietal lobule	−22	−60	54	1303	6.61	4.76	.00
WM	−22	−52	44		6.17	4.57	
L superior parietal lobule	−28	−60	46		5.85	4.42	
L middle occipital gyrus	−26	−94	14	239	6.38	4.66	.01
WM	−24	−78	16		4.15	3.48	
R superior occipital gyrus	26	−74	40	543	5.95	4.46	.00
WM	20	−58	42		4.78	3.86	
R superior occipital gyrus	24	−60	54		4.64	3.78	
R middle occipital gyrus	32	−88	16	323	5.56	4.27	.00
R middle occipital gyrus	34	−78	12		4.63	3.77	
WM	30	−84	6		4.34	3.60	
R middle temporal gyrus	62	−46	−8	240	4.96	3.96	.01
R middle temporal gyrus	70	−36	−8		4.89	3.92	
R inferior temporal gyrus	48	−54	−10		4.45	3.67	
L inferior occipital gyrus	−42	−66	−8	233	4.73	3.83	.01
L middle occipital gyrus	−44	−78	−2		4.45	3.67	
L inferior occipital gyrus	−36	−72	−4		4.18	3.50	

All results $p < .05$, FWE, at cluster level, and $p < .001$, uncorrected, at peak level. Minimum cluster size: $k = 50$. WM = white matter.

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

significance level ($p < .001$, uncorrected, $k > 50$). Similarly, no whole-brain differences were observed between voices and thoughts overall or between indirect forms of speech and thought, either at corrected or uncorrected levels.

DISCUSSION

The aim of this study was to explore further the effect of direct speech in the brains of readers. The main finding of our results was to replicate the original effect reported by Yao et al. (2011), namely, that direct speech in short stories is accompanied by elevated responses in voice-selective auditory regions of the brain, when compared with indirect speech. Our findings go further than those of Yao et al. in two key ways. First, by comparing direct and indirect references for speech and thoughts, our ROI results demonstrate a specific effect of reference for characters who are represented as speaking, but not when they are represented as thinking. Second, this direct speech effect appears to extend beyond voice-selective auditory cortex to also include regions that are used when making inferences about communicative intentions, based on a ToM localizer task (Ciaramidaro et al., 2007). This pattern of results, therefore, supports the earlier observation that readers spontaneously engage sensory cortices when faced with direct speech, but it also implicates higher-order processes associated with gauging character intention and meaning.

Evidence of a direct speech effect in auditory cortex is consistent with previous findings that such regions are recruited during silent reading of characters' speech (Yao et al., 2011, 2012), which is in turn suggestive of auditory verbal imagery being used during this process. This aligns with behavioral evidence of phonologically detailed imagery being involved in silent reading of various kinds (Kurby & Zacks, 2013; Filik & Barber, 2011). There is debate around how specific any such voice representation would be: Kurby, Magliano, and Rapp (2009) have argued that such effects are specific to familiar voices only, whereas Petkov and Belin (2013) propose that any kind of voice simulation is likely to reflect a generic speaking voice. Their argument for this is based on phonological information specific to voice identity usually being associated with anterior temporal cortex, whereas those associated with direct speech in Yao et al. (2011), for example, are more focused on posterior temporal regions (Petkov & Belin, 2013). Our findings cannot easily arbitrate between these two possibilities (general vs. specific voices), as voice-selective auditory regions were identified along the length of the superior temporal gyri bilaterally. However, we would speculate that any simulation of a generic or specific voice is likely to vary considerably across individuals. When asked, readers describe drawing upon a wide range of active and creative strategies to imagine the voices of characters, including other familiar voices and their own voice (Alderson-Day et al., 2017).

Perhaps more notable is the suggestion of direct speech effects also being present in cortical regions often associated with ToM in general and understanding others' intentions in particular.[4] We chose a localizer task that aimed to minimize superficial overlaps with the primary task—using cartoons instead of a written story format—and focused specifically on assessing understanding of communicative intentions over other types of ToM reasoning, such as inferring false beliefs (Ciaramidaro et al., 2007; Walter et al., 2004). This produced a network that, in our sample, primarily centered around bilateral TPJ regions but also included precuneus and mPFC in subsets of participants. Evidence of a direct speech effect in these regions provides at least prima facie support for the idea that text presented in this way prompts engagement with what a character intends to say (Yao et al., 2011, 2012), despite the mixed behavioral evidence that direct reference primes any further communicative information about characters (Eerland et al., 2013). Moreover, our analysis suggests involvement of these regions at a comparable level to responses in auditory networks, as indicated by the lack of any interaction effect across the two ROIs.

Drawing strong conclusions about the role of these regions in processing direct speech is fraught with difficulty. The areas highlighted by our ToM task are often implicated in a range of attentional and cognitive processes (Spreng et al., 2009; Mitchell, 2008), and making broader claims based on the prior literature raises the risk of reverse inference (Poldrack, 2006). Using Neurosynth (Yarkoni et al., 2011), which provides at least a systematic approach to informal reverse inference (Poldrack, 2011), the strongest responses in the localizer task were in two regions where the most common associations in the literature are with "mind tom" and "theory mind" (with posterior probabilities of .87–.90). Similarly, in the exploratory whole-brain analysis, the right MTG peak in particular showed high z scores for tests of association ($z = 12.00$) and uniformity ($z = 14.39$) in a ToM meta-analytic map of 181 studies (Yarkoni et al., 2011).

These regions have also been observed in similar work examining sociocognitive responses to fiction reading by Tamir, Bricker, Dodell-Feder, and Mitchell (2016), although in their study, they observed preferential engagement of the mPFC for social content in stories (describing a person's mental content), with medial temporal cortex more closely indexing story vividness. In contrast, most of our participants (15/16) activated the right TPJ on our localizer task (compared with only six for mPFC), and this was the only ToM region to be identified in our whole-brain analysis comparing direct and indirect speech. The right TPJ cluster that we observed in this analysis included peaks in the right AG, extending dorsally and caudally from areas that are often linked to representing others' mental states (Bzdok et al., 2013). Both left

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

and right AG have been associated with support for the default mode network, via the generation and processing of transmodal information in the absence of stimulus input (Murphy et al., 2018) and modality-independent contributions to imagery (Daselaar, Porat, Huijbers, & Pennartz, 2010). The right AG has also recently been implicated in making valence judgments from nonverbal cues: In a paradigm where participants were asked to judge the intentions of musical alien "signals," variations in the consonance and dissonance of the stimuli (roughly corresponding to positive and negative emotions) modulated this region specifically (Bravo et al., 2017). The broader extension of this cluster, therefore, may reflect the generation and maintenance of intention-related imagery, rather than representing characters' mental states, or social content more generally. This being associated with posterior ToM regions over mPFC would also be consistent with van Overwalle's (2009) distinction between a posterior ToM subsystem supporting representation of temporary and perceptually based intentions and goals, versus an anterior pFC system that tracks and integrates enduring social information over time.

When taken together, these findings broadly support the interpretation of direct speech made by Yao et al. (2011). Recall that, for Yao and colleagues, direct speech prompts auditory imagery as a means of modeling speaker prosody (and, ultimately, communicative intent). A counterhypothesis, provided by Eerland, Zwaan, and colleagues, is that direct reference acts primarily as a cue to simulate verbatim linguistic content—in other words, emphasizing the words but, arguably, not the speaker (Eerland & Zwaan, 2018; Eerland et al., 2013). Our data suggest that direct reference has a specific effect for speech, and this extends to regions that would be consistent with inferring communicative intentions. Moreover, this can be distinguished from the overall effect of direct reference, which primarily shows greater engagement in visual areas of occipital and parietal cortex (see Table 4C).

A curious characteristic of our data is the apparently contradictory results for a direct speech effect in auditory regions, which was evident in the ROI analysis, but not for the whole-brain contrast. This likely reflects (i) individual variability in the temporal voice area (Belin et al., 2000), (ii) the effect of the more conservative statistical correction required across the whole brain, and (iii) the fact that both direct and indirect speech activate a range of overlapping temporal regions, with any subsequent difference in beta values being likely to be subtle. Nevertheless, it should be noted that prominent differences across the cortex were observed in the right TPJ (as discussed), left SPL, and left IFG, much more obviously than for regions of the auditory cortex. The involvement of the latter in particular is consistent with greater demand being placed on inner speech production to support the representation of direct speech, given the common association of Broca's area with silent articulation (Alderson-Day & Fernyhough, 2015; Kühn, Fernyhough, Alderson-Day, & Hurlburt, 2014; Simons et al., 2010; Shergill et al., 2001). Evidence from psycholinguistics research suggests that greater involvement of articulatory processes in silent speech results in more detailed acoustic properties being represented in auditory imagery (Oppenheim & Dell, 2010), and both external and internal speech have been shown to consistently modulate auditory cortical responses (Okada, Matchin, & Hickok, 2018; Ylinen et al., 2014; Shergill et al., 2002). In addition, two recent studies of inner speech have highlighted how right-hemisphere homologs of left-hemisphere language regions are recruited when speech of others must be imagined (Grandchamp et al., 2019; Alderson-Day et al., 2016). A potential model, then, would be that a reader coming across direct speech in a text is prompted to generate a communicatively plausible perceptual simulation, via inner speech, which involves the left IFG and right TPJ working in concert to modulate voice-selective regions of the auditory cortex. This is not to suggest that inner speech (and other auditory imagery processes) would not be evidenced in each of the task conditions (given the widespread activation vs. baseline seen for all conditions; see Table 2 and Figure 4A) but rather that direct speech could place a specific demand on internal articulatory processes. In this scenario, direct reference effects in auditory cortex would plausibly not be the primary component of the reader's response but a secondary consequence of inner speech (and ToM) processes, which may explain their relative prominence in our whole-brain results.

Although the present results appear to have much to say about how speech is treated by readers, they perhaps say less about what is happening for characters' thoughts. Despite having received early theoretical attention in stylistics (Sharvit, 2008; Sotirova, 2004), until now, qualitative differences between direct and indirect modes of speech and thought representation have scarcely been empirically investigated (for some exceptions using free indirect discourse, see Fletcher & Monterosso, 2015; Bray, 2007). A plausible assumption would be that thought presentation (in direct or indirect reference) would be more likely to engage ToM resources—that is, a main effect of thoughts—compared with speech. Why, then, was this not seen? Insights from contemporary cognitive narratology may be useful here, particularly in relation to the problem of "accessibility" of others' thoughts. Ordinarily, stories that are used to assess ToM require the reader to make inferences about the mental states of others; their actual beliefs are not made explicit and may even conflict with the literal and immediate content of what they say and how they act (e.g., Saxe & Kanwisher, 2003). Fictional narratives may sometimes exploit this "accessibility gap" (e.g., a suspect in a mystery could have hidden motives), but they are also notable because they can give us apparent access to other minds via direct and indirect references (Bernini, 2016; Cohn,

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

1978). Although the stimuli used in our experiment contained mental content, they did not necessarily make demands in terms of mental state inference—in the thought trials used in our experiment, the inner life of the character is laid bare (e.g., "He thought that he should go to the shop"). Direct speech, in contrast, does not signal the intonation, emotion, or intention of a character—they must be simulated or otherwise inferred by the reader, in a way that the ToM system is often considered to do (Saxe & Kanwisher, 2003). As such, although counterintuitive, our findings are in line with common views about mental state inference (Spreng et al., 2009). It may also be the case that direct speech in general is more vivid and salient than direct reference to thinking, given the whole-brain differences between speaking and thinking seen in anterior insula and dorsal ACC (Uddin, 2016), and the quicker orienting times we observed for speech trials. Engagement with fictional storyworlds and characters is often argued to depend on the "experiential traces" the reader brings from his or her own life (Zwaan, 2008): The more we have access to an experience in the real world, the more it will be used to generate vivid and imaginative responses during reading. When one considers the diminished, quasi-perceptual phenomenology that verbal thoughts are often claimed to possess (Prinz, 2011; Jones & Fernyhough, 2007), it is perhaps no surprise that characters' thinking in a text did not provide distinct patterns of activation that were as distinct as for direct speech.

Another perspective—also provided by cognitive literary studies—is to consider how fictional minds may be differently represented from the outside and the inside. Kuzmičová (2013), for example, has suggested that we experience characters' speech in literary texts as either "outer reverberations" (when we read, as vicarious listeners, about a character overtly speaking) and "inner reverberations" (when we voice a character's words within his or her perspective). In parallel, Caracciolo (2014) has highlighted the contrast between attributing intentions to characters and the direct, inner enactment of a character's thoughts and fictional consciousnesses more broadly. These distinctions parallel the extensive literature on perspective-taking and how this is instantiated in the brain (e.g., Ruby & Decety, 2001). It could be the case that our different conditions prompted readers to adopt first- or third-person perspectives in response to speech compared with thoughts or direct compared with indirect reference. However, the direction of these shifts is not straightforward: Although it is sometimes assumed that direct speech necessarily prompts adopting a first-person perspective (speaking as the character), it is also understood as focusing the reader on what it would be like to hear the character speak to them (Clark & Gerrig, 1990). Similarly, thoughts could be seen to prime a first-person perspective (thinking "from the inside"), but this will likely depend on the position of the narrator, the reader's identification with the character, and the wider context of the narrative (Kuiken, Miall, & Sikora, 2004). As such, a key area for further exploration is to systematically examine how perspective shifts potentially interact with direct reference effects and speech/thought distinctions.

This study has a number of limitations. First, it was necessary to exclude some participants because of partial data from eye-tracking or either of the independent localizer tasks, limiting the overall sample size. This also further skewed our sex ratio, such that male participants are underrepresented in our eventual sample (as can often be the case for psychology studies recruited from university populations, e.g., Dickinson, Adelson, & Owen, 2012). Given the wide variability in individual differences for reading responses, we chose to deploy these measures to be as specific as possible about both participants' onset and offset times of reading target sentences and to allow for the use of individually specific cortical networks. This did not prohibit the recruitment of a larger sample than the original study we sought to replicate (Yao et al., 2011), but for a topic (imagery) with typically small effects and potentially large variation, replication in larger samples will be required for the exploration of individual differences in imagery production across different kinds of readers. Inner speech and imagery are highly susceptible to individual differences in day-to-day use (Alderson-Day et al., 2016), and effects of expertise (Borst, Niven, & Logie, 2011) and variation across readers seem highly likely.

Second, our use of direct reference for thoughts (such as *he thought "I should have finished this paper by now"*) could be questioned in terms of its relative familiarity for readers. One of our aims for the study was to use a stimulus that could act as a typographical and grammatical control comparison for direct and indirect speech. Although use of quotation marks for thoughts does feature in narratives, indirect references might be thought of as many authors' default option when referring to characters' mental states (Leech & Short, 2007). An alternative form of reference—such as using italics to mark characters' thoughts—may have been more familiar to readers but would also have added further typographical differences to the original contrast of interest: direct versus indirect speech. The lack of any behavioral differences (in terms of accuracy or reading time) between the thought conditions, and the lack of any pairwise or whole-brain differences, would suggest that this had little effect on our participants. However, further careful behavioral (and, arguably, interdisciplinary) work—incorporating the valuable insights of cognitive literary studies—is clearly required to elucidate how readers interpret these kinds of text constructions when depicting characters' mental states.

Finally, a related point about generalizability concerns the fictional stories used in the experiment. For experimental use, we used very minimal stories that were unlikely to prompt extensive use of many of the processes thought to be relevant to a reader's experience of a

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

text, whether that involves identification with characters, use of prior knowledge, management of expectations, or feelings of transportation (Miall, 2011; Green, 2004; Kuiken et al., 2004). As such, this is still a very artificial reading scenario for many participants. We cannot rule out the possibility that there was something about this situation in particular that may have posed unusual demands or biased readers' responses, such as encouraging them to pay attention to or engage more with specific aspects of the text (such as voices in particular). Our attentional checks would adjudicate against this interpretation—no significant differences in accuracy were observed across the various task conditions—but, in considering ecological validity, the experiential gap between full stories and these experimental sketches must be borne in mind.

Notwithstanding these limitations, our findings have important implications for future research on fiction, reading, and imagination more generally. Our data broadly support social cognitive approaches to fiction (Oatley, 2016; Tamir et al., 2016), but in a complex and unexpected way. On the one hand, the potential involvement of ToM in simulating episodes of characters' speech opens a new avenue for research on fiction and mentalizing; on the other hand, our findings for representing characters' thoughts challenge the idea that engaging with the mental states of others via fiction necessarily involves (or could even enhance) ToM processes. Our findings also highlight how readers likely draw on multiple perceptuomotor resources to support a socially informed simulation of speech, where prompted by the text. This is, arguably, a creative and constructive process on the part of the reader, which will be contingent on their own imaginative skills and experience. Along with comparing individual differences in this process, contrasting forms of reference for speech and verbal thoughts offer a comparative methodology for exploring how readers track speaking and thinking through more complicated narratives. Free indirect discourse, as seen in many modernist texts, demands that the reader follow closely, or even make their own inference, about exactly who is speaking or thinking in a story (see Waugh, 2011, for a discussion of this topic). Here, reference or its absence could be considered as an experimental tool to challenge the readers and place them in situations of uncertainty about the speech and thoughts in a narrative (as in Fletcher & Monterosso, 2015). In this respect, more challenging texts offer an opportunity to push at the limits of readers' creative and imaginative capacities.

Conclusions

In conclusion, references to direct speech in fictional stories are associated with the recruitment of not only voice-selective auditory cortex but also regions that may implicate gauging of characters' communicative intentions. Moreover, this is a process that is apparently specific to speech. We cannot conclude on the basis of these findings that the function of this process is communicative inference per se, but we speculate that it goes beyond a purely perceptual simulation of voice and requires coordination between inner speech and ToM resources. To experience a character's voice in a story, in this sense, may be about not only what they say but also how they say it and what they intend.

Acknowledgments

This work was supported by the Wellcome Trust (WT098455 and WT108720). John Foxwell, Lucy May, and Anthony Atkinson are thanked for their assistance with piloting and eye-tracking; and David Smailes and the Hearing the Voice team are thanked for their contributions to the early development of the research question. For more information on this process, please see Fernyhough (2015).

Reprint requests should be sent to Ben Alderson-Day, Department of Psychology, Durham University, Science Laboratories, South Road, Durham DH1 3LE, UK, or via e-mail: benjamin.alderson-day@durham.ac.uk.

Notes

1. Exposure to literary fiction in particular has also been proposed to enhance readers' ToM skills (Oatley, 2016; Kidd & Castano, 2013), although such claims have not always been supported in replication attempts (Kidd & Castano, 2018; Klein et al., 2018). For recent meta-analyses on this topic, see Dodell-Feder and Tamir (2018) and Mumper and Gerrig (2017).
2. A full list of the stories and questions used is available at community.dur.ac.uk/benjamin.alderson-day/RVT_full_stim_aldersonday.pdf.
3. This analysis was also run with ROIs that explicitly excluded areas identified in the auditory localizer task, leading to almost identical results: no main effects and a significant interaction, $F(1, 15) = 9.29$, $p = .008$, $\eta_p^2 = .38$. On an individual level, auditory and ToM ROIs overlapped in only 2 of 16 participants, and this was to a minimal degree.
4. Although Yao et al.'s (2011) main analysis was ROI driven, they also conducted an exploratory whole-brain analysis that primarily identified regions of posterior temporal cortex and occipital-fusiform regions associated with word reading.

REFERENCES

Alderson-Day, B., Bernini, M., & Fernyhough, C. (2017). Uncharted features and dynamics of reading: Voices, characters, and crossing of experiences. *Consciousness and Cognition*, *49*, 98–109.

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*, 931–965.

Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., & Fernyhough, C. (2016). The brain's conversation with itself: Neural substrates of dialogic inner speech. *Social Cognitive and Affective Neuroscience*, *11*, 110–120.

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 446–459.

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.
Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*, 359–375.
Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
Bernini, M. (2016). The opacity of fictional minds: Transparency, interpretive cognition and the exceptionality thesis. In P. Garratt (Ed.), *The cognitive humanities: Embodied mind in literature and culture* (pp. 35–54). London: Palgrave Macmillan.
Borst, G., Niven, E., & Logie, R. H. (2011). Visual mental image generation does not overlap with visual short-term memory: A dual-task interference study. *Memory & Cognition*, *40*, 360–372.
Bortolussi, M., & Dixon, P. (2003). *Psychonarratology: Foundations for the empirical study of literary response*. Cambridge: Cambridge University Press.
Bravo, F., Cross, I., Hawkins, S., Gonzalez, N., Docampo, J., Bruno, C., et al. (2017). Neural mechanisms underlying valence inferences to sound: The role of the right angular gyrus. *Neuropsychologia*, *102*, 144–162.
Bray, J. (2007). The 'dual voice' of free indirect discourse: A reading experiment. *Language and Literature: International Journal of Stylistics*, *16*, 37–52.
Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage*, *16*, S497.
Brück, C., Kreifelts, B., Gößling-Arnold, C., Wertheimer, J., & Wildgruber, D. (2014). 'Inner voices': The cerebral representation of emotional voice cues described in literary texts. *Social Cognitive and Affective Neuroscience*, *9*, 1819–1827.
Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., et al. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage*, *81*, 381–392.
Caracciolo, M. (2014). *The experientiality of narrative: An enactivist approach*. Boston: Walter de Gruyter.
Caracciolo, M., & Hurlburt, R. T. (2016). *A passion for specificity: Confronting inner experience in literature and science* (Forthcoming ed.). Columbus: The Ohio State University Press.
Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., et al. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia*, *45*, 3105–3113.
Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*, 764–805.
Cohen, J. (1969). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.
Cohn, D. (1978). *Transparent minds*. Princeton, NJ: Princeton University Press.
Coulmas, F. (2011). *Direct and indirect speech*. New York: Walter de Gruyter.
Daselaar, S. M., Porat, Y., Huijbers, W., & Pennartz, C. M. (2010). Modality-specific and modality-independent components of the human imagery system. *Neuroimage*, *52*, 677–685.
Dickinson, E. R., Adelson, J. L., & Owen, J. (2012). Gender balance, representativeness, and statistical power in sexuality research using undergraduate student samples. *Archives of Sexual Behavior*, *41*, 325–327.
Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. *Journal of Experimental Psychology: General*, *147*, 1713–1727.
Eerland, A., Engelen, J. A., & Zwaan, R. A. (2013). The influence of direct and indirect speech on mental representations. *PLoS One*, *8*, e65480.
Eerland, A., & Zwaan, R. A. (2018). The influence of direct and indirect speech on source memory. *Collabra: Psychology*, *4*, 5.
Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, *9*, 338.
Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review*, *28*, 225–262.
Fernyhough, C. (2015). The experimental design hackathon. In C. Fernyhough, A. Woods, & V. Patton (Eds.), *Working knowledge: Transferable methodology for interdisciplinary research. Hearing the voice*. Durham: Durham University.
Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PLoS One*, *6*, e25782.
Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128.
Fletcher, A., & Monterosso, J. (2015). The science of free-indirect discourse: An alternate cognitive effect. *Narrative*, *24*, 82–103.
Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven, CT: Yale University Press.
Gerrig, R. J., Brennan, S. E., & Ohaeri, J. O. (2001). What characters know: Projected knowledge and projected co-presence. *Journal of Memory and Language*, *44*, 81–95.
Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., et al. (2019). The ConDialInt model: Condensation, dialogality, and intentionality dimensions of inner speech within a hierarchical predictive control framework. *Frontiers in Psychology*, *10*, 2019.
Green, M. C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, *38*, 247–266.
Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, *14*, 311–327.
Herman, D. (2008). Narrative theory and the intentional stance. *Partial Answers: Journal of Literature and the History of Ideas*, *6*, 233–260.
Herman, D. (2013). *Storytelling and the sciences of mind*. Cambridge, MA: MIT Press.
Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, *16*, 391–399.
Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*, 377–380.
Kidd, D. C., & Castano, E. (2018). Reading literary fiction and theory of mind: Three preregistered replications and extensions of Kidd and Castano (2013). *Social Psychological and Personality Science*, *10*, 522–531.
Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., et al. (2018). Many labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*, *1*, 443–490.
Köder, F., Maier, E., & Hendriks, P. (2015). Perspective shift increases processing effort of pronouns: A comparison between direct and indirect speech. *Language, Cognition and Neuroscience*, *30*, 940–946.
Kühn, S., Fernyhough, C., Alderson-Day, B., & Hurlburt, R. T. (2014). Inner experience in the scanner: Can high fidelity apprehensions of inner experience be integrated with fMRI? *Frontiers in Psychology*, *5*, 1393.

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Kuiken, D., Miall, D. S., & Sikora, S. (2004). Forms of self-implication in literary reading. *Poetics Today*, *25*, 171–203.
Kukkonen, K. (2014). Presence and prediction: The embodied reader's cascades of cognition. *Style*, *48*, 367–384.
Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition*, *112*, 457–461.
Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, *126*, 338–349.
Kuzmičová, A. (2013). Outer vs. inner reverberations: Verbal auditory imagery and meaning-making in literary narrative. *Journal of Literary Theory*, *7*, 111–134.
Leech, G., & Short, M. (2007). *Style in fiction: A linguistic introduction to english fictional prose* (2nd ed.). New York: Routledge.
Magrassi, L., Aromataris, G., Cabrini, A., Annovazzi-Lodi, V., & Moro, A. (2015). Sound representation in higher language areas during language generation. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 1868–1873.
Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*, 173–192.
Maslej, M. M., Oatley, K., & Mar, R. A. (2017). Creating fictional characters: The role of experience, personality, and social processes. *Psychology of Aesthetics, Creativity, and the Arts*, *11*, 487–499.
Miall, D. S. (2011). Emotions and the structuring of narrative responses. *Poetics Today*, *32*, 323–348.
Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*, 262–271.
Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291.
Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *11*, 109–120.
Murphy, C., Jefferies, E., Rueschemeyer, S.-A., Sormaz, M., Wang, H., Margulies, D. S., et al. (2018). Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. *Neuroimage*, *171*, 393–401.
Oatley, K. (2011). *Such stuff as dreams*. Malden, MA: Wiley-Blackwell.
Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences*, *20*, 618–628.
Okada, K., Matchin, W., & Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic Bulletin & Review*, *25*, 423–430.
Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*, 1147–1160.
Palmer, A. (2004). *Fictional minds*. Lincoln, NE: University of Nebraska Press.
Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., et al. (2012). How silent is silent reading? Intracerebral evidence for top–down activation of temporal voice areas during reading. *Journal of Neuroscience*, *32*, 17554–17562.
Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239.
Petkov, C. I., & Belin, P. (2013). Silent reading: Does the brain 'hear' both speech and voices? *Current Biology*, *23*, R155–R156.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.
Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*, 692–697.
Polvinen, M. (2016). Enactive perception and fictional worlds. In P. Garratt (Ed.), *The cognitive humanities* (pp. 19–34). London: Palgrave Macmillan.
Prinz, J. (2011). The sensory basis of cognitive phenomenology. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 174–196). Oxford: Oxford University Press.
Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*, 135–147.
Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, *4*, 546–550.
Ryan, M.-L. (1999). Immersion vs. interactivity: Virtual reality and literary theory. *SubStance*, *28*, 110–137.
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., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692–699.
Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34.
Schurz, M., Tholen, M. G., Perner, J., Mars, R. B., & Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: A review using probabilistic atlases from different imaging modalities. *Human Brain Mapping*, *38*, 4788–4805.
Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy*, *31*, 353–395.
Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Jr., Murray, R. M., et al. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, *16*, 219–227.
Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, *31*, 241–253.
Simons, C. J., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., Dominguez, M. G., et al. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry*, *67*, 232–237.
Sotirova, V. (2004). Connectives in free indirect style: Continuity or shift? *Language and Literature: International Journal of Stylistics*, *13*, 216–234.
Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*, 489–510.
Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience*, *11*, 215–224.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.
Uddin, L. Q. (2016). *Salience network of the human brain*. San Diego, CA: Academic Press.
van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. San Diego, CA: Academic Press.

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*, 829–858.
Vilhauer, R. P. (2016). Inner reading voices: An overlooked form of inner speech. *Psychosis: Psychological, Social and Integrative Approaches*, *8*, 37–47.
Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, *16*, 1854–1863.
Waugh, P. (2011). Thinking in Literature: Modernism and contemporary neuroscience. In D. James (Ed.), *The legacies of modernism: Historicising postwar and contemporary fiction* (pp. 75–95). Cambridge: Cambridge University Press.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.
Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*, 3146–3152.
Yao, B., Belin, P., & Scheepers, C. (2012). Brain "talks over" boring quotes: Top–down activation of voice-selective areas while listening to monotonous direct speech quotations. *Neuroimage*, *60*, 1832–1842.
Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670.
Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., et al. (2014). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cerebral Cortex*, *25*, 1576–1586.
Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Columbus, OH: Ohio State University Press.
Zwaan, R. A. (2008). Experiential traces and mental simulations in language comprehension. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 165–180). Oxford: Oxford University Press.
Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*, 611–619.

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/32/9/1637/1862017/jocn_a_01571.pdf by guest on 05 May 2021

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn_a_01571 by guest on 28 March 2024