Social Cognitive and Affective Neuroscience Advance Access published October 6, 2015

OXFORD

Social Cognitive and Affective Neuroscience, 2015, 1–10

doi: 10.1093/scan/nsv114
Original article

Reading fiction and reading minds: the role of simulation in the default network

Diana I. Tamir,[1] Andrew B. Bricker,[2] David Dodell-Feder,[3] and Jason P. Mitchell[3]

[1]Department of Psychology, Princeton University, Princeton, NJ, USA, [2]Department of English, McGill University, Montreal, Quebec, Canada, and [3]Department of Psychology, Harvard University, Cambridge, MA, USA

Correspondence should be addressed to Diana I. Tamir, Peretsman-Scully Hall, Washington Street, Princeton, NJ 08544, USA.
E-mail: dtamir@princeton.edu

Abstract

Research in psychology has suggested that reading fiction can improve individuals' social-cognitive abilities. Findings from neuroscience show that reading and social cognition both recruit the default network, a network which is known to support our capacity to simulate hypothetical scenes, spaces and mental states. The current research tests the hypothesis that fiction reading enhances social cognition because it serves to exercise the default subnetwork involved in theory of mind. While undergoing functional neuroimaging, participants read literary passages that differed along two dimensions: (i) vivid *vs* abstract and (ii) social *vs* non-social. Analyses revealed distinct subnetworks of the default network respond to the two dimensions of interest: the medial temporal lobe subnetwork responded preferentially to vivid passages, with or without social content; the dorsomedial prefrontal cortex (dmPFC) subnetwork responded preferentially to passages with social and abstract content. Analyses also demonstrated that participants who read fiction most often also showed the strongest social cognition performance. Finally, mediation analysis showed that activity in the dmPFC subnetwork in response to the social content mediated this relation, suggesting that the simulation of social content in fiction plays a role in fiction's ability to enhance readers' social cognition.

Key words: fiction; reading; theory of mind; default network; simulation; functional MRI

Introduction

Readers of fiction can transcend the here-and-now to experience worlds, people and mental states that differ vastly from their local reality. The consequences of reading, however, extend far beyond the subjective experience of any one individual. Researchers from fields as diverse as evolutionary psychology, literary studies and anthropology have independently credited literacy as a possible explanation for such fundamental societal shifts as the decline in human violence over the past few centuries, the development of desire-based over rule-based social interactions, and the advent of 'modern subjectivity' (Lukacs, 1920; Watt, 1957; Ong, 1982; McKeon, 1987; Habermas, 1991; Pinker, 2011). Such large-scale societal impacts may nevertheless begin with small behavioral changes in individual readers, who demonstrate greater civic engagement, including higher levels of volunteering, donating and voting, than non-readers (Katz, 2006). How might reading effect its influence on these individuals?

Recent research in psychology suggests that readers make good citizens because reading may improve one's ability to empathize with and understand the thoughts and feelings of other people. Readers of fiction score higher on measures of empathy and theory of mind (ToM)—the ability to think about others' thoughts and feelings—than non-readers, even after controlling for age, gender, intelligence and personality factors (Mar *et al.*, 2006, 2009, 2010). Developmental work has likewise shown a correlation between reading and social cognition. Children between the ages of four and six who were exposed to more

Received: 25 May 2015; **Revised:** 1 September 2015; **Accepted:** 1 September 2015

juvenile fiction performed better at ToM tasks than children with less exposure, again controlling for such potentially confounding factors as age, gender, vocabulary and parental income (Mar *et al.*, 2010). Other developmental work has similarly demonstrated that the frequency of parent–child picture book reading and parents' use of mental state terms predict false-belief task performance (Adrian *et al.*, 2005), and that the use of stories that contain more emotional, social and psychologically convincing content predicts empathy and socioemotional adjustment (Aram and Aviram, 2009). Recent experimental research has further shown that fiction reading plays a causal rather than just correlational role in the development of social-cognitive skills, such that among adults, fiction reading enhances ToM performance (Kidd and Castano, 2013) and empathy (Bal and Veltkamp, 2013).

However, not all reading improves social cognition. One study found that after controlling for demographic factors, personality traits and exposure to non-fiction and other fiction genres, only exposure to Romance significantly predicted ToM performance (Fong *et al.*, 2013). In another series of studies, though high-quality 'literary' fiction consistently improved social cognition, lower-quality fiction and non-fiction did not (Kidd and Castano, 2013). Indeed, people who regularly read non-fiction do not have better social abilities and may have worse social abilities than more infrequent readers of non-fiction (Mar *et al.*, 2006, 2009). However, at least one study found that people randomly assigned to read either literary fiction or literary non-fiction did not differ in empathy change pre- to post-reading; only when taking participants' openness into account did the expected difference between fiction and non-fiction emerge (Djikic *et al.*, 2013).

Developmental research further suggests that quality and genre may not be the only features that moderate reading's ability to improve social cognition; content, and the kinds of cognitive demands that a piece makes on readers, may also play an important role. In one experiment, children who read books that required them to *construct* their own social interpretations performed better on social-cognition tasks than children exposed to stories that explicitly provided such metacognitive language (Peskin and Astington, 2004). In a similar vein, adults assigned to read fiction over a 1-week period demonstrated positive changes in empathy only when they reported high emotional transportation into the story (Bal and Veltkamp, 2013), suggesting that immersion into and simulation of the mental and emotional lives of the characters may be the mechanism of change.

Taken together, these findings suggest that the effectiveness with which literature improves social cognition may depend on how well it demands attention to others' mental states. That is, perhaps literary fiction improves social cognition to the extent that it requires readers to mentally construct social contexts. Such high-quality practice in simulation—or the capacity to experience realities outside of the 'here-and-now', including hypothetical events, distant worlds, and other people's subjective experience—then translates into real-world consequences for readers' social cognition (Zunshine, 2006).

Because we now know a great deal about the neural networks involved in such simulation processes, work in neuroimaging presents a unique way to test this prediction. In particular, our brain's default network, which comprises the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), posterior superior temporal sulcus (pSTS), temporal parietal junction (TPJ), anterior medial temporal gyrus and medial temporal lobes (MTLs), is responsible for supporting our capacity for simulation (Raichle *et al.*, 2001; Buckner and Carroll, 2007; Schacter and Addis, 2007; Spreng *et al.*, 2009). The default network is recruited whenever people conjure up experiences outside of their local experiences, such as thinking about the future or the past, mentally constructing places and spaces, imagining hypothetical events and thinking about another's perspective (Okuda *et al.*, 2003; Addis *et al.*, 2007; Buckner and Carroll, 2007; Hassabis *et al.*, 2007; Schacter *et al.*, 2007; Szpunar *et al.*, 2007; Botzung *et al.*, 2008; Hassabis and Maguire, 2009; Tamir and Mitchell, 2011). The entirety of the default network has been associated with simulation in general but research has also demonstrated that two distinct subnetworks of the default network are recruited differentially when simulating vivid spatial content and mental content, respectively (Andrews-Hanna *et al.*, 2010). More specifically, scene construction and the simulation of vivid physical spaces rely on neural structures within more ventral aspects of the default network, such as the ventromedial prefrontal cortex (vmPFC), hippocampal and parahippocampal gyri and retrosplenial cortex (Hassabis *et al.*, 2007; Hassabis and Maguire, 2009), structures that comprise the default network's MTL subnetwork. Conversely, studies of human social cognition suggest that thinking about people and mental states recruits a separate set of regions, including the dorsomedial prefrontal cortex (dmPFC), anterior temporal poles and TPJ (Mitchell *et al.*, 2002; Saxe and Wexler, 2005; Mitchell, 2008), structures that comprise the default mode's dmPFC subnetwork. Researchers have noted the overlap between the dmPFC subnetwork and the network of brain regions associated with ToM, suggesting that ToM may rely on simulation processes (Andrews-Hanna *et al.*, 2010; Spreng and Grady, 2010; Mars *et al.*, 2012).

Fiction reading, which engenders simulations of vivid and social content, also recruits the default network (Mar, 2004, 2011). Thus, the overlap between reading and simulation is perhaps unsurprising, given that narratives often invoke vivid descriptive language to transport readers to far-off places, and engage readers with characters' actions, interactions and mental states. Both the simulation of physical spaces and of mental entities provides a plausible explanation for why reading reliably activates the default network (Mar, 2004; Yarkoni *et al.*, 2008, 2011; Mason and Just, 2009; Speer *et al.*, 2009). However, the relation between default network activity and the simulation of scenes and minds, respectively, during fiction reading has yet to be explicitly tested. This study first empirically tests the hypothesis that reading recruits the default network because it evokes both of these types of simulation.

In addition, we capitalize on this hypothesized overlap between simulation and fiction reading to test the hypothesis that, by recruiting the default network while reading, readers may practice the types of simulation necessary for solving social tasks. However, not all types of simulation should play a commensurate role in the relation between reading and enhanced social cognition. As suggested by previous research (Peskin and Astington, 2004), only simulations of social content should provide relevant social practice, whereas simulations of non-social scenes, events or hypothetical scenarios should fail to provide relevant social practice. This proposed dynamic relationship between neural function and experience is supported by neuroplasticity literature, which has demonstrated that repeated engagement in cognitive processes can lead to positive changes in the neural networks supporting those cognitive processes (Draganski and May, 2008; Klingberg, 2010; Anguera *et al.*, 2013; Lovden *et al.*, 2013; Merzenich *et al.*, 2014). Thus, here, repeated engagement in social simulation *vis-à-vis* fiction reading may lead to beneficial changes in the default network, which may carry concomitant benefits for social ability. Said otherwise, fiction reading may impact social ability through its effect on the neural system supporting social simulation. We note the possibility that individuals who have greater ToM ability may read more fiction, in which case greater ToM ability may predispose individuals to engage in more simulation, which may make reading fiction more enjoyable. However, current empirical work suggests a causal effect of fiction reading on ToM (Peskin and Astington, 2004; Mar *et al.*, 2010; Bal and Veltkamp, 2013; Kidd and Castano, 2013), rather than the reverse. Thus, taken together, we further hypothesize that neural activity while simulating social content, but not vivid physical scenes, should mediate the relation between fiction reading and ToM.

To test these hypotheses, participants in this study underwent functional neuroimaging scanning while they read literary excerpts designed to engender both simulations of vivid physical scenes and simulations of social content. Outside of the scanner, participants provided measures of their reading behavior and ToM ability. We expect that reading, which engenders simulation, should preferentially recruit the default network. More specifically, simulations of vivid scenes should evoke activity in the MTL subnetwork of the default network whereas simulations of social content should evoke activity in the dmPFC subnetwork of the default network. Further, we expect that the extent of fiction reading should predict ToM performance, replicating previous research on the relation between reading and social cognition. Finally, to the extent that simulation of social content provides the practice necessary for improvements in social cognition, we expect that neural activity specific to social simulations should mediate this relation between fiction reading and ToM.

Method

Participants

Twenty-six (16 female) right-handed, native English speakers with no history of neurological problems participated in this study (*M* age = 21.2 years; range = 19–26 years). All participants provided consent in a manner approved by the Committee on the Use of Human Subjects at Harvard University.

fMRI reading task

While undergoing fMRI scanning, participants read a series of literary passages, excerpted from a wide variety of sources, including novels, biographies, magazines, newspapers and self-help books (Supplementary Materials A). Each passage came from a unique source. For each trial, participants were presented with one passage to read (*M* length = 85 words; range = 56–106). Instructions emphasized that participants should pay full attention as they read each passage and that they did not need to finish reading the passage within the time allotted. Passages remained on screen for up to 30 s. Participants pressed a button under their index finger if they finished the passage before 30 s after which the passage was replaced by a fixation cross for the remainder of the 30 s period. Four seconds of fixation followed each 30 s reading period.

Passages varied systematically along two orthogonal dimensions: (i) the vividness with which they described physical scenes (Vivid *vs* Abstract) and (ii) whether or not they described a person or a person's mental content (Social *vs* Non-social). Passages were selected and categorized based on the pretest ratings of a separate set of participants across a variety of features: Social passages were selected to be highly social and personal, and contain either one person or groups of people; Vivid passages were selected to be high on vividness and movement and low on abstractness. There were a total of four passage types. Vivid/Social passages describe vivid scenes or events that include references to mental states, individuals or groups of people. Vivid/Non-social passages describe vivid physical scenes or events but lack references to mental states, individuals or groups of people. Abstract/Social passages use abstract language and thus lack easily imagined physical scenes but include references to mental states, individuals or groups of people. Finally, Abstract/Non-social passages use abstract language and lack imaginable physical scenes, people and mental states. Importantly, the four types of passages did not differ in their pretest ratings of boringness or wordiness or in average reading time (Supplementary Materials B).

During scanning, participants read 13 passages of each type, for a total of 52 passages. In addition, 13 fixation periods (each lasting 30 s) were included. The 52 passages and fixation periods were presented in a random order for each participant, divided among five consecutive runs of 442 s each.

fMRI data acquisition and analysis

Functional data were acquired using a gradient-echo echo-planar pulse sequence (TR = 2 s; TE = 30 ms) on a 3T Siemens Trio. Images were acquired using 36 axial, interleaved slices with a thickness of 3 mm (0.5 mm skip) and 3×3 in-plane resolution and online motion correction. Functional images were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). Data were first spatially realigned to correct for head movement and then unwarped to reduce image distortions. Images were then normalized to a standard anatomical space (3 mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute). Normalized images were then spatially smoothed using an 8 mm FWHM Gaussian kernel.

Preprocessed images were analyzed using a general linear model in which the events were modeled using a canonical hemodynamic response function and covariates of no interest (session mean, linear trends and six motion realignment parameters). Events began at the onset of the presentation of the passage and lasted for a duration of either 30 s, or until the participant indicated that he or she had finished reading the passage by pressing a button. Trials were conditionalized based on the type of passage presented, resulting in four conditions of interest: Vivid/Social, Vivid/Non-social, Abstract/Social *and* Abstract/Non-social. To test whether reading recruited the default network, primary analyses identified voxels in which BOLD response differed along the two dimensions of interest; that is, vividness (Vivid/Social + Vivid/Non-social) > (Abstract/Social + Abstract/Non-social) and sociality (Vivid/Social + Abstract/Social) > (Vivid/Non—social). Analyses were performed individually for each participant, and contrast images generated within each participant were subsequently entered into a second-level analysis treating participants as a random effect. Group level whole-brain contrasts employed an experiment-wise threshold of $P < 0.05$ corrected for multiple comparisons per Slotnick and Schacter's (2004) specifications; Monte Carlo simulations indicated use of a statistical criterion of 54 or more contiguous voxels at a voxel-wise threshold of $P < 0.01$.

To test the hypothesis that the MTL subnetwork would preferentially respond to the vividness of passages and the dmPFC subnetwork to the socialness of passages, we assessed neural responses to the reading task in independently defined regions-of-interest (ROIs). ROIs were defined as 8 mm spheres centered on the coordinates for each of the 11 regions independently identified by Andrews-Hanna (2010). Using functional connectivity analyses, Andrews-Hanna *et al.* (2010) identified 11 default network regions, divided into three functionally and anatomically distinct subnetworks: (i) a MTL subnetwork that comprises the hippocampal formation, parahippocampal cortex, retrosplenial cortex, posterior intraparietal lobe and vmPFC; (ii) a dmPFC subnetwork that comprises the dmPFC, temporal pole, lateral temporal cortex and temporal-parietal junction; and (iii) a 'core' subnetwork that comprises the PCC and mPFC (Andrews-Hanna *et al.*, 2010). Parameter estimates for the Social > Non-social and Vivid > Abstract contrasts were extracted from these ROIs to examine how each subnetwork responds to simulating physical and mental events during reading. Activity within each subnetwork was then calculated as the average parameter among the regions composing that network. Follow-up analyses evaluated individual ROI response within each network. Four outliers (>2.5 s.d. of the mean) were identified and Winsorized by replacing them with the next highest non-outlying value and adding 10% of that value to maintain variance. Of note, the direction and significance of the findings did not change when using the Winsorized values.

Behavioral measures

In addition to the reading task, participants also completed behavioral surveys outside of the scanner. We measured participants' exposure to both fiction and non-fiction and their social-cognitive abilities to assess whether participants showed the expected relation between fiction reading and ToM (Mar *et al.*, 2006, 2009).

Fiction reading. Participants completed the author recognition test (ART) to assess the extent to which they read fiction and non-fiction in their daily lives. This measure was originally developed by Stanovich and West (1989) but participants in this study saw the most recently updated and validated version of the ART developed by Mar *et al.* (2006). For this test, participants were presented with the names of fiction authors (50 names), non-fiction authors (50 names) and 40 foils, and were asked to place checkmarks next to the names that they recognized as authors. Participants needed only to recognize a name as that of an author but were not required to have read any of the author's work. Participants were told that some of the names were of people who are not writers. In this way the ART discourages guessing and overcomes potential issues of self-report bias. Following Stanovich and West (1989), fiction and non-fiction ART scores were calculated separately as the number of fiction or non-fiction author names a participant recognized, respectively, minus the number of foils they reported recognizing.

ToM. Participants also completed a ToM task that assessed the extent to which they spontaneously think about intentions when judging an individual's behavior. For this task, participants read 48 vignettes in which an actor engages in a behavior with either a negative or neutral outcome, on the basis of either a negative or a neutral intention (Young *et al.*, 2010a). The negative and neutral intentions were fully counterbalanced with the negative and neutral outcomes across 48 stimuli, resulting in 12 vignettes of four types: (i) no harm—neutral intention/neutral outcome, (ii) intentional harm—negative intention/negative outcome, (iii) accidental harm—neutral intention/negative outcome and (iv) attempted harm—negative intention/neutral outcome. After reading each vignette, participants judged the permissibility of the actor's behavior on a scale from 1 (*forbidden*) to 5 (*permissible*). Participants read and answered all moral judgment questions at their own pace. Given that the intention differs from the outcome in the accidental and attempted harm scenarios, judgments of moral permissibility reflect the extent to which participants take into account the actor's intention as opposed to the outcome. Thus, to the extent that participants consider the actor's intention, participants should judge the action in accidental harm scenarios as *more* permissible and judge the action in attempted harm scenarios as *less* permissible (see Supplementary Materials C for more information). Using this task, previous researchers have shown that responses to these two scenario types provide a sensitive measure of ToM (Young *et al.*, 2010a,b; Moran *et al.*, 2011).

fMRI reading task memory assessment. Finally, participants completed a surprise memory assessment for the stimuli presented during the reading task. This task measured the extent to which they had remained attentive during the scanning session based on their ability to recognize a series of sentences that either had been presented during scanning ('old') or had not been seen ('new'). Performance was assessed by calculating *d*-prime for each participant (Supplementary Materials D).

Brain-behavior analysis

One outlier was identified in the moral judgment data and was Winsorized. We evaluated the relationship between the behavioral measures with bivariate Pearson correlations. These values are accompanied by bias-corrected and accelerated 95% CIs generated from 5000 bootstrap samples in SPSS. To evaluate the hypothesis that fiction reading impacts ToM ability, in part, through its effect on the neural bases of social simulation, we tested a mediation model with fiction reading (fiction ART score) as the predictor variable, neural activity for social simulation as the mediator and as an index of ToM ability, performance on the moral judgment task as the outcome variable. We used a non-parametric bootstrapping procedure to estimate the indirect effect; that is, the path from the predictor to the outcome variable through the mediator (fiction reading → neural basis of social simulation → ToM ability). Estimates of the indirect effect are accompanied by bias-corrected and accelerated 95% CIs derived from 5000 bootstrap samples. As measures of effect size, we provide the proportion of variance accounted for by the mediated effect (R^2_{med}) and, as recommended by Preacher and Kelley (2011), κ^2, which represents the ratio of indirect effect observed relative to the maximum possible indirect effect. This analysis was implemented in SPSS with the PROCESS macro (Hayes, 2013).

Results

fMRI results

Our primary analyses identified brain regions that responded to the two features of interest: (i) the vividness with which passages described physical scenes and events and (ii) whether or not they described a person or a person's mental content. Consistent with earlier research, both vivid passages and social passages recruited regions of the default network significantly more than abstract and non-social passages. A whole-brain random-effects contrast of Social>Non-social passages revealed activity in dmPFC, vmPFC, lateral temporal cortex from the temporal pole to the TPJ bilaterally, bilateral hippocampi and bilateral IFG (Figure 1A; Table 1). A whole-brain random-effects contrasts of Vivid>Abstract passages revealed robust activity in MTL structures, including hippocampus and parahippocampus bilaterally, retrosplenial cortex and precuneus (Figure 1B; Table 1).

To test the hypothesis that subnetworks of the default network respond differentially to each of these two features of the passages, we assessed neural responses to the four passage types within the three subnetworks identified by Andrews-Hanna *et al.* (2010) (Figure 1C). Consistent with the whole-brain analysis demonstrating that the subnetworks differentially respond to the social content and vividness of the passages, a 3 Subnetwork (Core, MTL, dmPFC) $\times$ 2 Vividness (Vivid, Abstract) $\times$ 2 Sociality (Social, Non-Social) repeated-measures ANOVA revealed an interaction between Subnetwork and Sociality, $F(2, 50) = 3.51$, $P = 0.04$, partial $\eta^2 = 0.12$, Subnetwork and Vividness, $F(2, 50) = 19.64$, $P < 0.001$, partial $\eta^2 = 0.44$, and a three-way interaction between Subnetwork, Sociality and Vividness, $F(2, 50) = 4.96$, $P = 0.01$, partial $\eta^2 = 0.17$.

Follow-up 2×2 repeated-measures ANOVA within subnetworks revealed that the vividness of the passages significantly affected activity in the MTL subnetwork, $F(1, 25) = 4.38$, $P < 0.05$, *Cohen's* $d = 0.42$, but the presence of people or mental content in the passages had no effect on the MTL subnetwork, $F(1, 25) = 0.21$, $P = 0.65$, $d = 0.09$. This suggests that the MTL network responded most robustly to passages that vividly described scenes but did not differentiate between passages with or without people. No interaction effects between the social and vivid factors were observed, $F(1, 25) = 2.66$, $P = 0.12$, partial $\eta^2 = 0.10$.

In contrast, a 2×2 repeated-measures ANOVA over activity in the dmPFC subnetwork revealed two main effects. First, the presence of people in the passages significantly affected activity in the dmPFC subnetwork, $F(1, 25) = 8.21$, $P < 0.01$, $d = 0.57$, such that this subnetwork responded more robustly to the presence of people and mental states in literary passages. Unexpectedly, we also observed a second main effect: abstract passages elicited more activity in the dmPFC subnetwork than vivid passages, $F(1, 25) = 28.72$, $P = 0.001$, $d = 1.07$, suggesting that the dmPFC subnetwork responds robustly to abstract content. No interaction effects between the social and vivid factors were observed, $F(1, 25) = 0.01$, $P = 0.92$, partial $\eta^2 = 0.00$.

The core subnetwork did not show differential responses to Social *vs* Non-social passages, $F(1, 25) = 0.48$, $P = 0.49$, $d = 0.14$ or Vivid *vs* Abstract passages, $F(1, 25) = 1.50$, $P = 0.23$, $d = -0.24$. No interaction effects between the social and vivid factors were observed, $F(1, 25) = 0.18$, $P = 0.67$, partial $\eta^2 = 0.01$. Responses in individual regions within the network are presented in Supplementary Materials E.

Thus, both the whole-brain and ROI analyses suggest that the default network does indeed respond differentially to literary passages depending on their content. Different subnetworks of the default network distinguished between the vividness of a passage and the social content of a passage. The MTL network responded most robustly to passages designed to be easy to simulate because they are rich in vivid details, whereas the dmPFC subnetwork responds most robustly to passages designed to be easy to simulate because they contain references to people or mental states.

Behavioral results

Performance on the scanner task and behavioral measures are presented in Supplementary Materials D and F, respectively. We found that people who read more fiction were more likely to take intentions into account when judging attempted harm scenarios (i.e. negative intention/neutral outcomes). Specifically, participants' fiction ART scores were significantly correlated with their ratings on the moral judgment task, $r(24) = -0.44$, $P = 0.02$, 95% CI [−0.66, −0.12] (Figure 2C), such that greater fiction reading was associated with judging actions as *less* permissible on attempted harm scenarios.

This positive association between reading and ToM was specific to fiction reading. Non-fiction ART scores among participants did not correlate with moral judgments of failed harm, $r(24) = -0.15$, $P = 0.48$, 95% CI [−0.48, 0.22], even though the extent to which participants read fiction and non-fiction was highly correlated, $r(24) = 0.80$, $P < 0.001$, 95% CI [0.64, 0.91]. Such findings replicate numerous previous studies that demonstrate that exposure to fiction, but not non-fiction, predicts enhanced ToM (Mar *et al.*, 2006, 2009, 2010; Kidd and Castano, 2013).

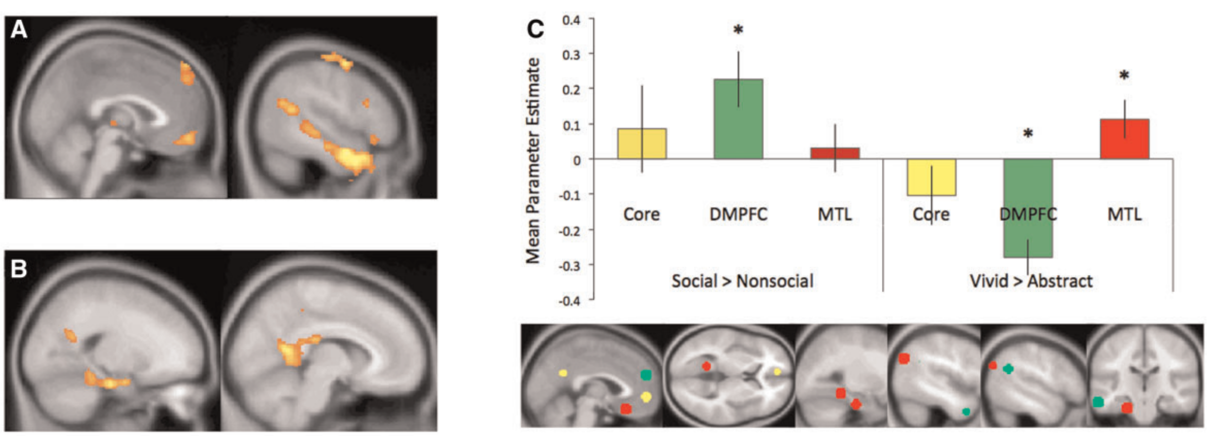

Fig 1. BOLD differences for main effect of (A) *Social > Non-social* (B) *Vivid > Abstract*, (C) and results of ROI analysis of both contrasts. Both whole-brain and ROI analyses show that the dmPFC subnetwork of the default network responded most robustly to literary passages containing people or mental content, whereas the MTL subnetwork responded most robustly to vivid physical descriptions.

Table 1. Peak voxel and cluster size for all regions obtained from a contrast of *Social* > *Non-social* and *Vivid* > *Abstract* (cluster-level corrected $P < 0.05$)

Anatomic label	x	y	z	Volume	Max t
Social > non-social					
Anterior temporal pole	34	31	−42	5362	6.99
	−44	9	−36	6189	6.96
Primary motor cortex	−42	−15	72	1159	6.94
	54	1	58	591	5.74
	22	−5	84	202	4.49
Inferior frontal gyrus	60	27	6	502	5.21
	46	29	−10	88	3.39
Cerebellum	26	−83	−34	586	5.01
	−26	−79	−38	146	4.21
	0	−51	−36	77	3.12
Dorsomedial prefrontal cortex	−10	59	46	1741	4.75
Ventromedial prefrontal cortex	6	49	−14	302	4.38
Occipital cortex	16	−107	24	55	3.98
	−8	−81	10	139	3.57
	−50	−97	6	55	3.43
	16	−79	14	61	2.86
Non-social > social					
Inferior temporal gyrus	−58	−55	−12	2554	7.45
Middle frontal gyrus	48	37	24	16836	7.02
	−30	35	−12		6.85
Inferior parietal lobule	−50	−49	52	12634	6.78
	52	−41	50		6.10
Middle temporal gyrus	60	−47	−10	1028	6.08
Superior temporal gyrus	52	1	−6	1186	4.68
	−52	−7	−2	142	4.29
Parahippocampal gyrus	24	−37	−4	113	4.40
	−32	−43	−10	589	3.85
	34	−21	−32	60	3.26
Cerebellum	56	−67	−42	348	4.14
Insula	−36	13	4	451	3.98
Occipital cortex	44	−75	4	98	3.48
Vivid > abstract					
Parahippocampal gyrus	18	−13	−20	1038	6.82
Inferior parietal lobule	−66	−35	38	1098	6.71
	62	−39	44	58	3.01
Fusiform gyrus	−34	−35	−20	1556	6.63
Angular gyrus	−36	−87	36	809	6.37
	44	−75	32	640	4.77
Middle temporal gyrus	−52	−63	−2	679	5.54
PCC	10	−53	14	1743	4.94
Middle frontal gyrus	−38	35	18	931	4.88
	48	51	24	66	3.21
Inferior frontal gyrus	48	31	6	326	3.98
	22	27	−12	162	3.38
Precuneus	−8	−35	46	378	3.60
STS	−40	−1	−20	87	3.29
Superior frontal gyrus	22	13	48	76	3.01
Abstract > vivid					
Superior temporal gyrus	−50	11	−24	10531	9.10
Inferior frontal gyrus	−50	27	−12		8.25
	64	19	22	61	3.36
	62	9	38	166	3.26
Middle temporal gyrus	−60	−23	−6	10531	7.96
	64	−41	−12	2546	6.31
Dorsomedial prefrontal cortex	−10	49	46	4539	8.08
Ventromedial prefrontal cortex	0	43	−20	1478	6.71
Temporo-parietal junction	−52	−63	30	1929	6.70
Insula	−32	3	8	591	6.54

continued

Table 1. (continued)

Anatomic label	x	y	z	Volume	Max t
Cerebellum	30	−83	−26	2659	6.12
	−42	−73	−22	167	3.99
Middle frontal gyrus	−46	11	52	1294	5.78
	40	59	−4	814	4.22
Inferior parietal lobule	44	−57	40	1132	4.69
Occipital cortex	46	−83	−4	272	4.43
	−4	−91	−20	71	3.51
	36	−83	18	78	3.40
PCC	−2	−43	32	191	3.90

Brain-behavior results

Based on the extant literature, we hypothesized that the neural basis of social simulation would explain the link between fiction reading and ToM ability. That is, fiction reading improves ToM, in part, through its effect on the neural basis of social simulation. Because reading mental and vivid physical passages differentially recruited distinct subnetworks of the default network, we were able to address this question using dmPFC subnetwork activity to social passages as an index of social simulation. If fiction reading enhances ToM because doing so activates or trains the neural networks involved in ToM (i.e. the dmPFC subnetwork), then we would expect the dmPFC subnetwork response to the *mental* simulations to mediate the relation between fiction reading and ToM.

These questions were addressed with mediation analysis. The paths between the predictor (fiction ART), mediator (dmPFC subnetwork activity for Social > Non-social) and outcome variable (moral judgments on the attempted harm scenarios) were significant in the predicted directions (Figure 2): fiction reading was positively associated with considering actors' intention when making moral judgments, and with dmPFC subnetwork activity; dmPFC subnetwork activity was positively associated with considering actors' intention when making moral judgments. Importantly, the direct effect of fiction reading on moral judgments was no longer significant when controlling for dmPFC subnetwork activity. Bootstrap analysis of the indirect effect (coefficient = −0.02, SE = 0.01) generated a CI that did not encompass zero, 95% CI [−0.05, −0.001], indicating that dmPFC subnetwork activity mediated the relationship between fiction reading and moral judgments ($R^2_{med} = 0.16$, 95% CI [0.02, 0.37]; $\kappa^2 = 0.21$, 95% CI [0.03, 0.46]).

Since the dmPFC subsystem also responded preferentially to abstract *vs* vivid passages, one possibility is that simulation of abstract and non-social features of fiction in this subsystem contributed to the mediated effect. We evaluated this possibility by running an additional mediation model controlling for the dmPFC subnetwork's response to Abstract > Vivid passages. The findings remain unchanged (Supplementary Materials G).

Another possibility is that fiction reading may impact ToM through its effect on the neural system selective for non-social simulation of vivid scenes. To evaluate this idea, we tested one additional model using neural activity in the MTL subnetwork for Vivid > Abstract as the mediator. Bootstrap analysis of the indirect effect revealed that non-social simulation of vivid scenes also did not mediate the relation between fiction and ToM ability (Supplementary Materials H).

Discussion

The link between fiction reading and ToM occurs at multiple levels of analysis. Psychologically, fiction readers possess

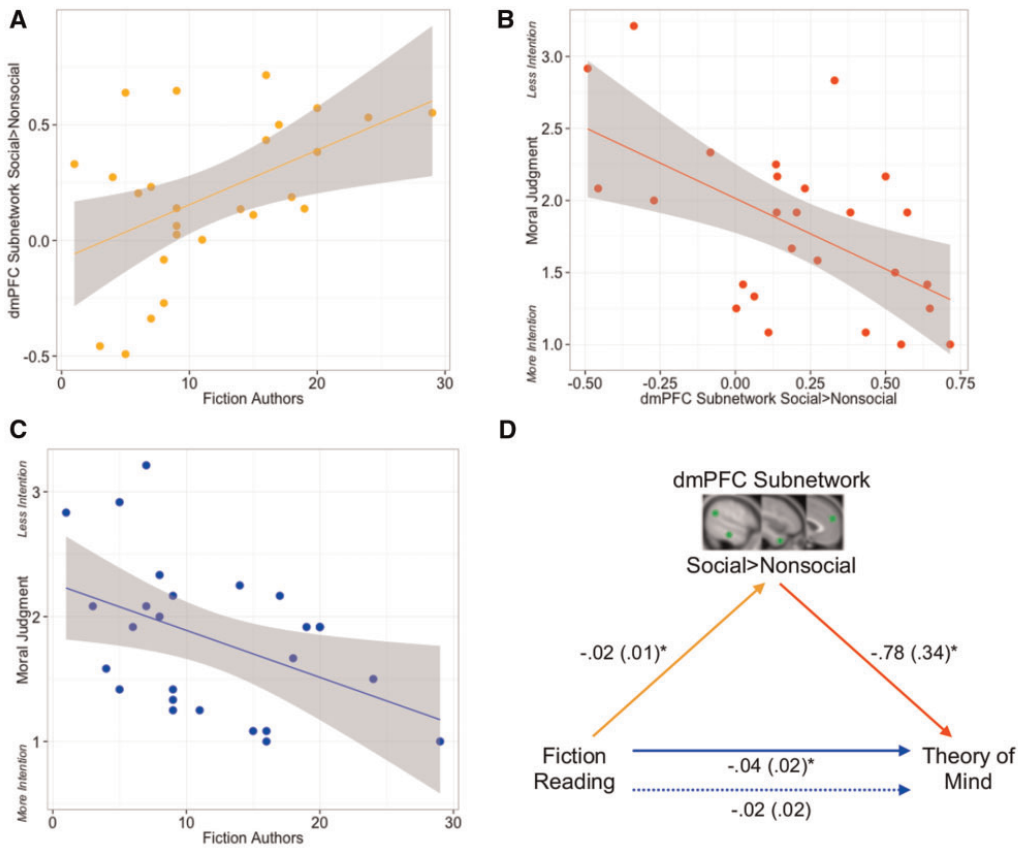

Fig 2. Depiction of significant correlations between (A) fiction reading scores on the fiction ART and dmPFC subnetwork activity during *Social > Non-social* passages, (B) between dmPFC subnetwork activity during *Social > Non-social* passages default activity and ToM task performance and (C) between fiction reading and ToM task performance. (D) The effect of fiction reading on ToM task performance through dmPFC subnetwork activity for social passages. Bootstrap analysis of the indirect effect indicated that dmPFC subnetwork activity mediated the relationship between fiction reading and ToM task performance. Unstandardized path coefficients shown with SE in parentheses for each path. The dotted line represents the direct effect of fiction reading on ToM task performance (i.e. controlling for the effect of dmPFC subnetwork activity). Note that the 'More Intention' and 'Less Intention' anchors on plots B and C are for visualization purposes only; participants rated each story on the Moral Judgment Task from 1 (*forbidden*) to 5 (*permissible*). *$P < 0.05$.

stronger social-cognitive abilities than both non-readers and non-fiction readers (Mar *et al.*, 2006, 2009, 2010). Historically, highly literate societies, especially societies that produced psychologically rich literature, function more empathically and less violently than less literate societies (Lukacs, 1920; Watt, 1957; Ong, 1982; McKeon, 1987; Habermas, 1991; Pinker, 2011). And neurally, fiction reading and social cognition recruit an overlapping neural network (i.e. the default network) (Mar, 2004, 2011). This study not only replicates previous findings that fiction reading both enhances social cognition and recruits the default network but also draws together these findings to test two hypotheses about the nature of this relation between reading fiction and ToM.

First, this study demonstrates that fiction reading recruits the default network because it elicits at least two distinct types of simulation: the simulation of vivid physical scenes and the simulation of people and minds. Each type of simulation recruited distinct subnetworks of the default network. Consistent with prior work evaluating non-social *vs* social scene construction (Hassabis *et al.*, 2014), simulations of physical scenes primarily recruited the MTL subnetwork of the default network, while simulations of people and minds primarily recruited the dmPFC subnetwork of the default network. Interestingly, and unexpectedly, this study also found that the dmPFC subnetwork was significantly more responsive to abstract content than vivid physical content. This finding may be consistent with prior literature on semantic and conceptual processing. For example, prior work has found dmPFC and left TPJ to be preferentially engaged during abstract or high-level construal tasks (e.g. generating semantic categories) *vs* low-level construal tasks (e.g. describing visual characteristics) regardless of the social content (Baetens *et al.*, 2014), a finding that has been further substantiated with meta-analytic data (Binder *et al.*, 2009). Similarly, focusing on the abstract features of personal memories preferentially recruits dmPFC and left TPJ, whereas focusing on concrete features of personal memories recruits aspects of the MTL subsystem (D'Argembeau *et al.*, 2014). Thus, our findings regarding the preferential response of the dmPFC subnetwork to abstract *vs* vivid passages converge with other findings on the role of this network in abstract processing.

Second, this study capitalized on these neural findings to test the hypothesis that fiction reading improves ToM by providing readers with the opportunity to exercise or practice mental simulation capacities that are also recruited during social-cognitive tasks. Said otherwise, fiction reading may impact ToM through its influence on the neural basis of social simulation. Mediation analysis was consistent with this idea. Specifically, we found that dmPFC subnetwork response to simulating people and minds mediated the relation between fiction reading and ToM. This effect was not changed when controlling for the dmPFC subnetwork's response to non-social abstract information. Furthermore, using MTL subnetwork activity to vivid *vs* abstract scenes, we further ruled out the possibility that fiction reading impacts ToM through its effect on non-social simulation. Together, the results suggest that any positive effect of reading fiction on social-cognitive abilities might be due to the influence of reading on neural networks involved in simulating social content and not non-social vivid scenes. This finding is consistent with that of behavioral work demonstrating that fiction reading over a 1-week period was associated with an increase in empathy only when readers reported a high level of transportation (i.e. simulation) of the characters' mental lives and story events (Bal and Veltkamp, 2013). We note that research in this field is nascent, and as such alternative models may be viable. For example, a reverse mediation model whereby ToM causally impacts the amount of fiction reading through its impact on social simulation. However, the model tested here, in which fiction reading impacts ToM through its effect on social simulation specifically, is most consistent with the extant literature regarding the nature and direction of the relations between the variables.

These findings also suggest that future research should focus on the *content* of literature to understand the relation between reading and ToM. Literature that effectively engages a reader in social content should be most likely to improve ToM; literature that does not successfully engage a reader in social thought, or that taxes a reader's imagination only with hypothetical events and places, should not. Previous researchers have studied how genre (fiction *vs* non-fiction) or the quality (literary *vs* non-literary) of such works improves ToM (e.g. Fong *et al.*, 2013; Kidd and Castano, 2013). The current research manipulated content irrespective of genre. As such, these findings suggest a need to reinterpret previous findings in terms of content, and the kinds of cognitive demands that content makes on readers. For example, literary fiction may just more effectively depict social content than low-quality literature, and fiction may more often traffic in social content than non-fiction.

Nevertheless, the fact that the social content of a passage may play an important role in shaping social cognition raises important consequences for future research in both psychology and literary studies alike. For instance, our study does not discretely define and test every kind of social or mental interaction. That is, our 'social' passages sometimes contain either one person or groups of people; they depict either the appearance of an individual or describe a character's abstract mental content; or they describe purely social interactions among groups of individuals. For both the psychologist and the literary scholar, a closer analysis of the content of the 'social' passages might reveal which aspects of social interaction and mental content fiction readers respond to most robustly, both neurally and psychologically. Future investigations into reading and social cognition would benefit significantly from a more in-depth understanding of which specific aspects of this range of social content most effectively drive the relation between fiction reading and enhanced ToM.

In a similar way, our passages range over a host of broad literary techniques associated with fiction without studying which techniques in particular provide the most relevant opportunities for simulation. Our 'social' passages include different techniques to supply readers with psychological information about their characters: direct representations of mental content; multi-level psychological inferences ('she believed that he believed ...'); free-indirect discourse (or the rendering of first-person thoughts into third-person narration); and physical actions linked to psychological states, thoughts and emotions (e.g. facial expressions and postures). Further narrative elements, such as issues related to first- and third-person narration, also present opportunities for further research into the relation among fiction, social-cognitive abilities and simulation.

The current findings must be interpreted in the context of several limitations. For one, given our sample size, the brain-behavior correlations and mediation findings may be overestimates of the true population effect (Button *et al.*, 2013). Additionally, though it is tempting to draw a causal inference from the mediation findings, the current data are cross-sectional and cannot definitively speak to a causal relationship between the variables. As such, future longitudinal research should endeavor to establish that fiction reading enhances social-cognitive abilities. Such a connection would hold extremely important real-world implications, perhaps guiding the direction of higher education and social initiatives more broadly, as well as potentially providing a tolerable and cost-effective intervention for social cognitive deficits in clinical populations. However, this is not to preclude the possibility that future research might establish an inverse relation: that social-cognitive abilities instead cause people to read fiction. In either case, there is still a great deal to be learned about the nature of social cognition, its relation to fiction reading, and their impacts on both personal choices and behavior.

Acknowledgments

The authors thank John Bender, Susanna Carmona, Juan Manuel Contreras, Eshin Jolly, Joe Moran, Brandi Newell, Kenneth Parreno, Amitai Shenhav, Emma Templeton, Blakey Vermeule, Adam Waytz and Jamil Zaki for helpful advice and assistance.

Funding

Andrew Bricker was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and the Andrew W. Mellon Foundation.

Supplementary data

Supplementary data are available at *SCAN* online.

Conflict of interest. None declared.

References

Addis, D. R., Wong, A. T., Schacter, D. L. (2007) Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, **45**(7), 1363–77.

Adrian, J. E., Clemente, R. A., Villanueva, L., Rieffe, C. (2005) Parent-child picture-book reading, mothers' mental state language and children's theory of mind. *Journal of Child Language*, **32**(3), 673–86.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., Buckner, R. L. (2010) Functional-anatomic fractionation of the brain's default network. *Neuron*, **65**(4), 550–62.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., *et al.* (2013) Video game training enhances cognitive control in older adults. *Nature*, **501**(7465), 97–101.

Aram, D., Aviram, S. (2009) Mothers' storybook reading and kindergartners' socioemotional and literacy development. *Reading Psychology*, **30**(2), 175–94.

Baetens, K., Ma, N., Steen, J., Van Overwalle, F. (2014) Involvement of the mentalizing network in social and non-social high construal. *Social cognitive and Affective Neuroscience*, **9**(6), 817–24.

Bal, P. M., Veltkamp, M. (2013) How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS One*, **8**(1), e55341.

Binder, J. R., Desai, R. H., Graves, W. W., Conant, L. L. (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, **19**(12), 2767–96.

Botzung, A., Denkova, E., Manning, L. (2008) Experiencing past and future events: functional neuroimaging evidence on the neural basis of mental time travel. *Brain and Cognition*, **66**, 202–12.

Buckner, R. L., Carroll, D. C. (2007) Self-projection and the brain. *Trends in Cognitive Sciences*, **11**(2), 49–57.

Button, K. S., Ioannidis, J. P., Mokrysz, C., *et al.* (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, **14**(5), 365–76.

D'Argembeau, A., Cassol, H., Phillips, C., Balteau, E., Salmon, E., Van der Linden, M. (2014) Brains creating stories of selves: the neural basis of autobiographical reasoning. *Social cognitive and Affective Neuroscience*, **9**(5), 646–52.

Djikic, M., Oatley, K., Moldoveanu, M. C. (2013) Reading other minds: effects of literature on empathy. *Scientific Study of Literature*, **3**(1), 28–47.

Draganski, B., May, A. (2008) Training-induced structural changes in the adult human brain. *Behavioural Brain Research*, **192**(1), 137–42.

Fong, K., Mullin, J. B., Mar, R. A. (2013) What you read matters: the role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, **7**(4), 370.

Habermas, J. (1991) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge: MIT Press.

Hassabis, D., Kumaran, D., Maguire, E. A. (2007) Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, **27**(52), 14365.

Hassabis, D., Maguire, E. A. (2009) The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**(1521), 1263.

Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., Schacter, D. L. (2014) Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cerebral Cortex*, **24**(8), 1979–87.

Hayes, A. F. (2013) *Introduction to Mediation, Moderation, and Conditional Process Analysis : A Regression-based Approach*. New York: The Guilford Press.

Katz, J. (2006) *Arts and Civic Engagement: Involved in Arts, Involved in Life*. Washington DC: National Endowment for the Arts.

Kidd, D. C., Castano, E. (2013) Reading literary fiction improves theory of mind. *Science*, **342**(6156), 377–80.

Klingberg, T. (2010) Training and plasticity of working memory. *Trends in Cognitive Sciences*, **14**(7), 317–24.

Lovden, M., Wenger, E., Martensson, J., Lindenberger, U., Backman, L. (2013) Structural brain plasticity in adult learning and development. *Neuroscience and Biobehavioral Reviews*, **37**(9 Pt B), 2296–10.

Lukacs, G. (1920) *The Theory of the Novel*. Cambridge: MIT Press.

Mar, R. A. (2004) The neuropsychology of narrative: story comprehension, story production and their interrelation. *Neuropsychologia*, **42**(10), 1414–34.

Mar, R. A. (2011) The neural basis of social cognition and story comprehension. *Annual Review of Psychology*, **62**(1), 103–34.

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., Peterson, J. B. (2006) Bookworms versus nerds: exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*, **40**(5), 694–712.

Mar, R. A., Oatley, K., Peterson, J. B. (2009) Exploring the link between reading fiction and empathy: ruling out individual differences and examining outcomes. *Communications*, **34**(4), 407–28.

Mar, R. A., Tackett, J. L., Moore, C. (2010) Exposure to media and theory-of-mind development in preschoolers. *Cognitive Development*, **25**(1), 69–78.

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., Rushworth, M. F. (2012) On the relationship between the "default mode network" and the "social brain". *Frontiers in Human Neuroscience*, **6**, 189.

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), 157–74.

McKeon, M. (1987) *The Origins of the English Novel, 1660-1740*. Baltimore: Johns Hopkins University Press.

Merzenich, M. M., Van Vleet, T. M., Nahum, M. (2014) Brain plasticity-based therapeutics. *Frontiers in Human Neuroscience*, **8**, 385.

Mitchell, J. P. (2008) Contributions of functional neuroimaging to the study of social cognition. *Current Directions in Psychological Science*, **17**(2), 142–6.

Mitchell, J. P., Heatherton, T. F., Macrae, C. N. (2002) Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, **99**, 15238–43.

Moran, J. M., Young, L. L., Saxe, R., *et al.* (2011) Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences*, **108**(7), 2688.

Okuda, J., Fujii, T., Ohtake, H., *et al.* (2003) Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. *Neuroimage*, **19**, 1369–80.

Ong, W. (1982) *Orality and Literacy: The Technologizing of the Word*. New York: Methuen & Co.

Peskin, J., Astington, J. W. (2004) The effects of adding metacognitive language to story texts. *Cognitive Development*, **19**(2), 253–73.

Pinker, S. (2011) *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking Adult.

Preacher, K. J., Kelley, K. (2011) Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological Methods*, **16**(2), 93–115.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., *et al.* (2001) A default mode of brain function. *Proceedings of the National Academy of Sciences*, **98**, 676–82.

Saxe, R., Wexler, A. (2005) Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, **43**(10), 1391–9.

Schacter, D. L., Addis, D. R. (2007) Constructive memory: the ghosts of past and future. *Nature*, **445**(7123), 27.

Schacter, D. L., Addis, D. R., Buckner, R. L. (2007) Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, **8**(9), 657–61.

Slotnick, S. D., Schacter, D. L. (2004) A sensory signature that distinguishes true from false memories. *Nature Neuroscience*, **7**, 664–72.

Speer, N. K., Reynolds, J. R., Swallow, K. M., Zacks, J. M. (2009) Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, **20**(8), 989–99.

Spreng, R. N., Grady, C. L. (2010) Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience*, **22**(6), 1112–23.

Spreng, R. N., Mar, R. A., Kim, A. S. N. (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**(3), 489–510.

Stanovich, K. E., West, R. F. (1989) Exposure to print and orthographic processing. *Reading Research Quarterly*, **24**(4), 402–33.

Szpunar, K. K., Watson, J. M., McDermott, K. B. (2007) Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, **104**, 642–7.

Tamir, D. I., Mitchell, J. P. (2011) The default network distinguishes construals of proximal versus distal events. *Journal of Cognitive Neuroscience*, **23**(10), 2945–55.

Watt, I. (1957) *The Rise of the Novel*. Berkeley: University of California Press.

Yarkoni, T., Speer, N. K., Balota, D. A., McAvoy, M. P., Zacks, J. M. (2008) Pictures of a thousand words: investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *Neuroimage*, **42**(2), 973–87.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., Damasio, A. (2010a) Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, **65**(6), 845–51.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., Saxe, R. (2010b) Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, **107**(15), 6753.

Zunshine, L. (2006) *Why We Read Fiction: Theory of Mind and the Novel*. Columbus: Ohio State University Press.