Poetics 52 (2015) 32–43

Contents lists available at ScienceDirect

Poetics

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

The effects of reading material on social and non-social cognition

Jessica E. Black *, Jennifer L. Barnes

University of Oklahoma, United States

ARTICLE INFO

Article history:
Received 10 February 2015
Received in revised form 30 June 2015
Accepted 2 July 2015
Available online

Keywords:
Fiction
Cognitive ability
Theory of mind

ABSTRACT

The purpose of this study was to investigate the effects of reading material on both social and non-social cognition. Prior research supports the hypothesis that reading fiction improves theory of mind (Kidd & Castano, 2013; Mar, Oatley, Hirsh, de la Paz, & Peterson, 2006; Mar, Oatley, & Peterson, 2009); however, little has been done to test its effects on other cognitive abilities. In this study, we tested the effect of reading literary fiction vs. non-fiction on both theory of mind and intuitive physics understanding. In line with previous research, results indicate a small but significant within-subject effect of reading material on theory of mind once other variables are controlled. Although the experimental manipulation (literary fiction vs. nonfiction) had no effect on intuitive physics understanding, we found that familiarity with fiction predicted intuitive physics ability. These results are discussed in terms of theories of fiction.

Reading for pleasure has been associated with better reading and writing performance in school, improved second language acquisition, increased understanding of other cultures and general knowledge, and greater community participation (Clark & Rumbold, 2006). People have speculated on the function of fiction at least since Aristotle (350 BCE/1902), but only in the last few decades has fiction been an object of research in psychology. What does reading *do* that makes it both enjoyable and beneficial? One answer to this question involves the effect of reading on theory of mind (cf. Djikic, Oatley, & Moldoveanu, 2013; Fong, Mullin, & Mar, 2013; Kidd & Castano, 2013; Mar, Oatley, Hirsh, de la Paz, & Peterson, 2006; Mar, Oatley, & Peterson, 2009). Theory of mind (ToM) is the ability to infer and reason about mental states: our own and other people's beliefs, desires, and intentions (Malle, 2006). Also referred to as *mind reading* (e.g., Zunshine, 2003), ToM demands that we understand that other people are agents, with both intentions and emotions that we can interpret and use to predict their behavior (Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001; Kidd & Castano, 2013).

Recent experimental research suggests that reading literary fiction (compared to nonfiction and popular fiction) may indeed enhance ToM (Kidd & Castano, 2013). These findings are in line with prior correlational research reporting that people's exposure to fiction, but not to nonfiction, was associated with increased scores on a ToM task (Djikic et al., 2013; Mar et al., 2006; Mar, Oatley, et al., 2009; Mar, Tackett, et al., 2009). Similarly, it has been shown that children exposed to more fictional media outperform their same-age peers on ToM tests (Mar, Tackett, & Moore, 2009). Research with people with autism spectrum conditions (ASC) gives convergent evidence: people with high-functioning autism seem to prefer nonfiction to fiction (Barnes, 2012), and do worse on ToM tasks compared with neurotypical controls matched by age and

* Corresponding author at: Department of Psychology, University of Oklahoma, Dale Hall Tower, Room 705, 455 W. Lindsey, Norman, OK 73071, United States. Tel.: +1 4053254511.
E-mail address: Jessica.black@ou.edu (J.E. Black).

http://dx.doi.org/10.1016/j.poetic.2015.07.001
0304-422X/

intelligence (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, Spong, et al., 2001). At the same time, individuals with ASC do better than neurotypical controls at tasks that depend on understanding systems and physical causality (Baron-Cohen, Wheelwright, Spong, et al., 2001; Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003; Baron-Cohen, 2010).

It has been suggested that the ability to understand people may compete in some way with the understanding of physical systems (Baron-Cohen et al., 2003; Baron-Cohen, 2009, 2010). Baron-Cohen proposed an explanation for autism that highlights the spectrum of individual differences that exists in both the ability and desire to understand the design and function of physical, rule-based systems and the ability and desire to understand the emotions, thoughts, and intentions of other people (Baron-Cohen et al., 2003). Individuals on or close to the autism spectrum tend to self-report higher scores on the former and lower scores on the latter when compared to neurotypical controls (e.g., Baron-Cohen & Wheelwright, 2004; Baron-Cohen, 2010; Goldenfeld, Baron-Cohen, & Wheelwright, 2005; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004; Wheelwright et al., 2006). Although these two types of abilities—understanding and responding to the emotions and thoughts of others and constructing and reasoning about physical, rules-based systems—are not necessarily related to each other, people who are higher in one tend to be lower on the other. Cross-culturally, women self-report, on average, a stronger affinity for understanding and responding to mental states and emotions than men do, whereas men score higher than women on a self-reported interest in and ability to reason about non-social systems (Wakabayashi et al., 2007; Wright & Skagerberg, 2012).

Although much of the research contrasting these two types of abilities uses self-reported measures (see Baron-Cohen, 2010 for an overview), performance-based tests have also been developed to examine the ability to read emotions and the ability to reason about physical objects. For example, the Reading the Mind in the Eyes test (RME; Baron-Cohen, Wheelwright, Hill, et al., 2001) shows participants a series of photographs of different people's eyes and asks them to identify the emotion depicted in each photograph. In contrast, the Intuitive Physics Test (IP; Baron-Cohen, Wheelwright, Spong, et al., 2001) measures participants' ability to reason about rules-based systems by asking them to indicate their understanding of physical causality based on a series of drawings. Prior research has shown that children with Asperger Syndrome do better on the IP and worse on the RME than neurotypical controls (Baron-Cohen, Wheelwright, Spong, et al., 2001). Although both the RME and the IP were originally developed for use with individuals on the autism spectrum, subsequent research has shown significant individual differences in performance on both tests within neurotypical populations as well (e.g. Brosnan, Gwilliam, & Walker, 2012; Wakabayashi, Sasaki, & Ogawa, 2012).

It is for this reason that the RME has often been used as a measure of ToM in research on fiction (e.g., Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013; Mar et al., 2006; Mar, Oatley, et al., 2009; Mar, Tackett, et al., 2009). A history of reading fiction, as assessed by familiarity with author names, has frequently been associated with higher scores on the RME (e.g., Djikic et al., 2013; Kidd & Castano, 2013; Mar et al., 2006), accounting for variability in scores on the RME over and above openness, age, experience, and fantasizing (Mar, Oatley, et al., 2009; Mar, Tackett, et al., 2009). Kidd and Castano carried out an experimental study in which participants were randomly assign to read either literary fiction or nonfiction, and then took the RME: participants scored higher on the RME after reading literary fiction than nonfiction, even after controlling for past exposure to fiction. In four subsequent studies, the effect of reading literary fiction was compared to the effect of reading popular fiction, and it was found that people consistently perform better on the RME after reading literary rather than popular fiction.

Although Kidd and Castano's (2013) experimental study provided support for the idea that reading literary fiction facilitates social cognition, the researchers did not investigate the effect of reading literary fiction on any non-social cognitive tasks. One possible interpretation of these results is that reading literary fiction facilitates our understanding of other people's minds per se. However, it is also possible that reading literary fiction may facilitate performance on a wider variety of cognitive tasks, including, but not limited to, those involving social cognition. In contrast, given the body of autism research reviewed above, it also seems relevant to ask whether reading literary fiction may improve our ability to read emotions at the expense of our ability to reason about physical systems.

Prior research has not investigated the effect of reading literary fiction (versus nonfiction) on performance on non-social cognitive tasks, such as the Intuitive Physics Test. Two possibilities exist: reading literary fiction may only affect ToM, or it may affect both ToM and the ability to reason about physical objects. If reading literary fiction does affect the ability to reason about objects, it could do so in one of two ways. It could be that the improvement in ToM ability seen after reading literary fiction corresponds to a decrease in the intuitive understanding of physical systems, in line with theories that contrast these two suites of abilities or the idea that reading fiction may flip participants into "people mode." Alternatively, reading literary fiction could improve *both* abilities. Under this view, the association of reading literary fiction with improved ToM could be evidence of the contribution of reading fiction to overall cognitive ability. For instance, perhaps fiction—particularly challenging fiction, such as the literary fiction found to increase ToM performance in Kidd and Castano (2013)—cues critical thinking, problem solving, and intellectual engagement and results in facilitating subsequent cognitive performance, whether the subsequent task is social or not.

The purpose of this research was to replicate and extend the research of Kidd and Castano (2013). We chose to focus on the effects of reading literary fiction versus nonfiction for two reasons. First, Kidd and Castano showed, in four experiments, that people tend to score higher on the RME after reading literary rather than popular fiction, but only one of their five experiments examined the effects of reading literary fiction versus nonfiction. Thus, because a goal of this research was to replicate the original findings, we focused on the condition (literary fiction versus nonfiction) that was not extensively

replicated within the original research. Secondly, we chose to focus on the effects of reading nonfiction because it seemed possible that reading nonfiction—particularly the type of nonfiction used by Kidd and Castano, which focuses on non-social content—might be associated with an increased ability to reason about physical systems; people with autism spectrum conditions both prefer non-social non-fiction (Barnes, 2012) and do better on tasks involving physical reasoning (Baron-Cohen, 2010).

The current research investigated the effects of reading literary fiction versus non-fiction on two tests: one that asks participants to read emotions (the Reading the Mind in the Eyes Test) and one that asks participants to reason about physical causality (the Intuitive Physics Test). Does reading fiction only affect ToM? In this case, we would expect to replicate past research and find higher scores on the RME after reading literary fiction rather than nonfiction, but find no difference for scores on the IP. Alternatively, does reading fiction affect non-social as well as social cognition? If so, does it facilitate or impede reasoning about rules-based systems?

We chose to investigate this question using a within-subjects design for two reasons. First, the strongest predictor of performance in the Kidd and Castano (2013) study we set out to replicate was prior exposure to fiction (as measured by scores on an Author Recognition Task [Acheson, Wells, & MacDonald, 2008]). Given that the effect of condition in the original experiment was somewhat small ($f = .23$), we chose to use a within-group design where lifetime exposure to fiction would be held constant across the literary fiction/nonfiction manipulation. A second reason for this decision was that we had no basis for estimating the size of any effect that reading literary fiction *or* nonfiction might have on performance on our measure of physical reasoning. Because within-subjects designs tend to be more sensitive than between-subjects designs, we thought that fully controlling for individual difference variables might increase the chances of revealing even a small effect of reading on physical reasoning. Accordingly, each participant was asked to read four short pieces (two literary fiction, two nonfiction), and each reading experience was followed with a test designed to examine either the ability to read emotions (the Reading the Mind in the Eyes Test) or the ability to reason about physical objects (the Intuitive Physics Test). Because participants took each of the two tests twice—once after reading literary fiction and once after reading nonfiction—we were able to directly compare their performance across the two conditions.

1. Method

1.1. Participants

Ninety-one undergraduates completed the study. Following Kidd and Castano (2013), we discarded cases with insufficient reading time (less than 1.5 *SD* below the mean; we chose 1.5 as a cut-off because two standard deviations would have placed it near zero—four seconds in one case); we also excluded those who spent less than 2 *SD* below the mean time on the RME. The final sample consisted of 60 students (80% female, mean age = 19.8) who completed the study in exchange for course credit.

1.2. Materials and instrumentation

1.2.1. Reading stimuli

Each participant read four pieces: two literary short stories and two nonfiction articles. The short stories and articles were chosen according to the criteria laid out by Kidd and Castano (2013). For both the literary fiction and the nonfiction conditions, one of the stimuli was taken directly from Kidd and Castano's (2013) materials, while the other was a new narrative, chosen based on their criteria. Both literary short stories used—"The Runner" by Don Delillo (1988), used in the Kidd & Castano paper, and "Puppy" by George Saunders (2007), chosen for this experiment—were written by contemporary award-winning authors, focused on a minimum of two characters, and were between 2000 and 3500 words in length. Two nonfiction articles of similar lengths were chosen from among recent online articles published in *Smithsonian Magazine*. Both articles focused on non-social content. The first, used by Kidd & Castano, "The Story of the Most Common Bird In the World" was about sparrows (Dunn, 2012), while the second, "The CIA'S Most Highly-Trained Spies Weren't Even Human," was about animals trained to spy (Vanderbilt, 2013). As in the experiment we were replicating, the stimuli chosen for the fiction and nonfiction conditions varied not only on the basis of fictionality—that is, whether or not they were true—but also on the basis of content (social versus non-social), narrativity (presenting a sequence of connected events, see Koopman & Hakemulder, 2015) and whether they were published in a literary (*Harper*'s and *The New Yorker*) or popular (*Smithsonian Magazine*) outlet.

1.2.2. Reading the mind in the eyes test

The RME (Baron-Cohen, Wheelwright, Hill, et al., 2001) has been used frequently as a measure of ToM in fiction research (e.g., Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013; Mar et al., 2006; Mar, Oatley, et al., 2009; Mar, Tackett, et al., 2009). The RME is a multiple-choice test that presents participants with photographs of the eye region of the face and requires them to choose which of four words best describes the emotion that the person in the photograph is feeling. For example, in one item, participants are shown a picture of a man's eyes and asked to choose whether that man feels *jealous, panicked, arrogant*, or *hateful.* Another item presents participants with a picture of a woman's eyes and asks whether she is *aghast, fantasizing, impatient*, or *alarmed.* For each item, participants get one point if they choose the correct target emotion and zero points if they do not. Points are totaled across thirty-six items for a final score ranging from 0 to 36.

1.2.3. Intuitive physics test

The IP (Baron-Cohen, Wheelwright, Spong, et al., 2001) was chosen as a measure of non-social cognition because it is a multiple choice test that focuses on physical reasoning and has been used in conjunction with the RME (e.g., Krajmer, Janosikova, Špajdel, & Ostatnikova, 2010; Wakabayashi et al., 2012). The IP presents participants with 20 drawings of physical objects and asks questions about how mechanisms and simple physics (weight, shape, etc.) work. For example, in one item, participants are shown a system of pulleys and are asked which direction one wheel rotates when another wheel is turned. Another item shows participants four blocks (a pyramid, a cone, a cylinder, and a rectangular cube) and asks which of the four would be the most difficult to push over, if all four weighed the same amount. Other items in the test are framed in more practical terms. For instance, one item presents participants with a sketch of a boat and asks which of five positions would be the best place to attach a rope to move the boat in a given direction; another asks which of three tank shapes would result in the water contained in the tank cooling the fastest. As with the RME, participants receive one point for every item on which they select the correct answer, for a total score ranging from 0 to 20.

1.2.4. The author recognition test

Participants were administered the Author Recognition Test (ART), developed by Acheson and colleagues (2008). This test, a modernization of Stanovich and West (1989) instrument, uses recognition of author names as a proxy for print exposure. Although Acheson and colleagues developed two tests, the Magazine Recognition Test, and the Author Recognition Test, geared at examining print exposure to books, following Kidd and Castano (2013), we issued only the Author Recognition Test, which Kidd and Castano used as "an index of general exposure to fiction" (378). In it, participants are shown 130 names; half of them correspond to real authors of fictional books, published across a variety of genres, and half are foils. Participants are told to select all the names with which they are familiar, and warned that there are fake names. Scores are calculated by subtracting the number of foils chosen from the number of real authors chosen.

1.2.5. The transportation scale

The TS (Green & Brock, 2000) is a 12-item self-report questionnaire that gauges the extent to which readers become involved with or "transported into" a story. Participants are asked to indicate their agreement on a 7-point Likert-type scale with statements such as "I had a vivid mental image of the characters" and "After the narrative ended, I found it easy to put it out of my mind" (reverse-scored). In this study, internal consistency reliability after reading fiction was $r_{\alpha} = .86$; after nonfiction it was $r_{\alpha} = .79$.

1.2.6. The short autism spectrum quotient

The AQ-S (Allison, Auyeung, & Baron-Cohen, 2012) is a ten-item brief form of the original 50-item AQ (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Participants are asked to indicate their agreement with items such as "I find it difficult to work out people's intentions" and "I often notice small sounds when others do not" on a four-point Likert-type scale. One point is be assigned for an answer of definitely or slightly agree on questions scored in the direction of higher AQ-S; on reverse-scored items, one point is assigned for an answer of definitely or slightly disagree. Allison et al. provide a cut-off point of six, with sensitivity of .88 and specificity of .91.

1.3. Procedure

Participants came into the lab and completed a computer survey via Qualtrics. The testing block was broken into three sections: (1) Reading and Testing Session 1, (2) Surveys and Demographics, and (3) Reading and Testing Session 2.

1.3.1. Reading and testing session 1

Participants were randomly assigned to read one of the four narrative stimuli described above and were told that upon completion, they would be expected to write a brief summary of what they had read. Because we were implementing a within-subject design and participants would ultimately be expected to read four separate pieces, this manipulation check was included both to ensure that participants had indeed read the assigned texts and to encourage them to continue reading each new text as the experiment proceeded. After having read the first text, participants were presented with either the Reading the Eyes in the Mind Test or the Intuitive Physics Test. They then proceeded to read a second narrative, matched in type to the first (e.g., if participants had been assigned to read "The Runner," they then read "Puppy"; if they had been assigned to read one of the two nonfiction articles, they then read the other nonfiction article). After reading the second text, participants took whichever of the two tests (RME and IP) they had not already taken. Finally, participants were asked to fill out the Transportation Scale to report their overall transportation into both of the narratives they had read.

1.3.2. Surveys and demographics

Before proceeding onto the next part of the experiment, participants were given a ten-minute break, during which they were instructed to leave the lab and walk around. Upon their return, they completed the ART and the AQ-S, as well as two other brief distractor surveys. They also reported demographic information (age and gender) and answered questions about their reading habits and preferences. After completing these measures, participants were then allowed to take another five-minute break.

1.3.3. Reading and testing session 2

After returning from the break, participants were randomly assigned to read one of the two texts they had not already received. Participants who read literary fiction during Session 1 read nonfiction during Session 2, and vice versa. The second session then proceeded in an identical fashion to the first session, with participants taking the RME after reading one of the texts and the IP after reading the other. For each text, participants were again asked to summarize what they had read. After reading both texts and taking both tests, participants filled out the Transportation Scale, reporting their transportation into the two narratives presented in Session 2.

1.4. Data analyses

Because our experimental design was within-subjects repeated measures, we used a multilevel, or mixed, model to test our hypotheses. Using mixed models to analyze repeated measures data avoids the problems inherent in using conventional ANOVA, and permits the researcher to include random variables, allowing for heterogeneity of variances, covariance, and individual differences at the subject level (Howell, 2010; Quene & van den Bergh, 2004) The SAS (version 9.3) PROC MIXED procedure was used to analyze the data (see Singer, 1998). We planned to enter scores on the RME as the dependent variable in the MODEL statement, with reading material (fiction vs. nonfiction), gender, ART, AQ-S, and TS as between subject fixed variables. Because preliminary analyses revealed order effects, we added order (fiction first vs, non-fiction first) and its interaction with reading material. The REPEATED statement was used to model the covariance matrix structure, with subject ID nested within order. Because fit indices were compared after the addition of each fixed variable, maximum likelihood estimation was used (Demidenko, 2004). Effect sizes for related measures were calculated according to Dunlap, Cortina, Vaslow and Burke (1996) formula for d that takes into account the correlation between repeated measures ($d = t\sqrt{2(1-r)/N}$). In line with our directional hypotheses based on Kidd and Castano (2013), as well as a wealth of correlational data suggesting that fiction is associated with increased performance on the RME (e.g. Fong et al., 2013; Mar et al., 2006; Mar, Oatley, et al., 2009), one-tailed tests were used to compare RME scores across the literary fiction and nonfiction conditions. To compare models, we used an adaptation of Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC; Schwarz, 1978) that includes a correction for sample size and model complexity.

2. Results

2.1. Preliminary analyses

First, the data were inspected to verify participant engagement in the tasks. As detailed in the methods section, all participants whose reading time on any of the four texts was less than 1.5 standard deviations below the mean were discarded. Analysis of the summary paragraphs written by participants confirmed that participants below this cut off had not read the material, as these summaries were either non-existent, grossly inaccurate, or failed to summarize anything that happened beyond the first few sentences of the short stories/articles they had read.

The grand mean for RME ($M = 26.11$, $SD = 4.33$; see Table 1 for descriptive statistics) was consistent with that reported by Baron-Cohen and colleagues (2001) for general population controls ($M = 26.2$, $SD = 3.6$), but lower than the mean they reported for undergraduates ($M = 28.0$, $SD = 3.5$). The distribution of scores after reading both fiction and nonfiction was normal. Men ($M = 26.08$, $SD = 4.75$) had similar scores to women ($M = 26.11$, $SD = 4.25$), Cohen's $d = .01$, overall, but gender was initially included in the model to test for interactions and model fit. Unsurprisingly, scores on the Transportation Scale suggested that participants were more engaged with fiction ($M = 46.8$, $SD = 11.9$) than with nonfiction ($M = 41.0$, $SD = 10.0$), $t(59) = 3.51$, $p = .001$, $d = 0.69$. Participants identified a mean of 9.0 ($SD = 6.2$, range = 34) authors on the ART; they had a mean of 0.8 misses (selecting a name that did not correspond to an author; $SD = 1.7$, range = 11). Final ART scores (calculated by subtracting misses from hits) had a mean of 8.5 ($SD = 6.05$, range = 35), and were positively skewed: a square root transformation was used for final analyses.

Table 1
Means and standard deviations for RME after reading fiction and nonfiction, the IP after reading fiction and nonfiction, the TS for fiction and nonfiction the ART, and the AQ-S.

	RME: Fiction	RME: Nonfiction	IP: Fiction	IP: Nonfiction	TS: Fiction	TS: Nonfiction	ART	AQ-S
Total	26.16	25.63	10.73	10.83	46.80	40.97	8.50	3.42
	(4.88)	(5.42)	(3.33)	(3.33)	(11.85)	(10.07)	(6.04)	(1.70)
Women	26.21	26.02	8.98	9.17	48.44	40.27	8.10	3.40
	(4.96)	(4.57)	(3.25)	(5.27)	(12.78)	(11.23)	(6.66)	(2.01)
Men	26.33	25.83	12.33	12.08	40.25	43.75	10.08	3.50
	(10.53)	(11.15)	(4.10)	(3.17)	(26.26)	(22.31)	(14.41)	(2.94)

Note. RME: the Reading the Mind in the Eyes test (Baron-Cohen, Wheelwright, & Hill et al., 2001); IP: Intuitive Physics test (Baron-Cohen, Wheelwright, & Spong et al., 2001); ART: Author Recognition Task (Acheson et al., 2008), AQ-S: Short Autism Quotient (Allison et al., 2012); TS: Transportation Scale (Green & Brock, 2000). Final model least squares means presented for RME and IP totals. Gender differences were significant for both IP tests and TS—Nonfiction.

Table 2
Zero-order correlations amongst scores on the RME and the IP after reading fiction and nonfiction, the ART, the AQ-S, and the TS for fiction and nonfiction.

	2	3	4	5	6	7	8
1. RME – Fiction	.74*** [.58, .84]	.22 [−.003, .43]	.20 [−.03, .43]	.27* [.02, .49]	−.22 [−.45, .03]	.13 [−.06, .31]	.07 [−.19, .31]
2. RME – Nonfiction	–	.21 [−.06, .44]	.27* [.01, .50]	.40** [.16, .59]	−.28* [−.51, −.01]	.16 [−.03, .34]	.05 [−.31, .22]
3. IP – Fiction		–	.84*** [.74, .91]	.35** [.11, .54]	−.27* [−.52, −.01]	−.10 [−.37, .18]	.16 [−.09, .39]
4. IP – Nonfiction			–	.30* [.04, .51]	−.27* [−.49, −.01]	−.03 [−.30, .25]	.16 [−.12, .40]
5. ART				–	−.35** [−.56, −.09]	.17 [−.05, .37]	.22 [−.08, .48]
6. AQ-S					–	−.13 [−.33, .06]	−.21 [−.43, .03]
7. TS – Fiction						–	.32* [.06, .53]
8. TS – Nonfiction							–

Note. $N = 60$. 95% bias corrected and accelerated confidence intervals ($N = 5000$ samples) presented below the correlations. RME: the Reading the Mind in the Eyes test (Baron-Cohen, Wheelwright, & Hill et al., 2001); IP: Intuitive Physics test (Baron-Cohen, Wheelwright, & Spong et al., 2001); ART: Author Recognition Task (Acheson et al., 2008), AQ-S: Short Autism Quotient (Allison et al., 2012); TS: Transportation Scale (Green & Brock, 2000).
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Scores on the ART were significantly and positively correlated ($.27 \leq r \leq .40$) with all other variables except transportation (see Table 2 for all zero-order relationships). Scores on the AQ-S were negatively correlated with all variables except transportation; the zero-order correlation between it and scores on the RME was stronger after reading nonfiction, $r(58) = -.28$, $p = .028$ than it was after reading fiction, $r(58) = -.22$, $p = .095$. Initial analyses suggested potential order effects: people tended to do better the second time they took the RME, and this effect appeared to be stronger if they had read literary fiction during Session 1 and nonfiction during Session 2 (mean gain = 1.10, $d = .26$) than if they had read the nonfiction articles first (mean gain = .32, $d = .07$). In other words, participants tended to score higher on the RME after reading nonfiction if they had already read a literary short story and then taken the RME than if the nonfiction article was the first thing they read (see Fig. 1). None of these differences were statistically significant, but given the small sample size, we entered order into the model and tested for an interaction once all covariates were included, using a mixed model design with order as a fixed variable (coded as 0 if they received the nonfiction block first and 1 if they received the fiction block first).

Scores on the IP were normally distributed after reading both fiction and nonfiction (grand mean = 9.70, $SD = 2.98$). Men ($M = 12.21$, $SD = 2.06$) scored higher than women ($M = 9.07$, $SD = 2.84$), $t(118) = 5.08$, $p < .001$, $d = 1.26$. There was no evidence of order effects. The mean difference between fiction and nonfiction IP scores overall was small ($M = 0.10$, $SD = 1.69$, $p = .649$, $d = .03$). Although the simple paired t-test for RME had also revealed a nonsignificant effect, it was more than twice as large (mean difference = 0.25, $d = .06$), even before accounting for order effects. These initial findings suggested that while reading material may have an effect on RME scores, it did not affect intuitive physics ability. Nonetheless, we proceeded to analyze the effect of reading material on RME and IP scores separately, using the same analyses, to confirm these results.

2.2. Theory of Mind (RME)

We began with a simple model, with order, reading material (fiction vs. non-fiction), and their interaction predicting scores on the RME, using an unstructured covariance matrix and testing the effect of reading material given order. Once order had been taken into account, reading fiction was associated with higher scores on the RME than reading nonfiction (mean difference = 0.935, $t(58) = 1.71$, $p_{\text{one-tailed}} = .046$, $d = 0.16$; see Table 3 for detailed model comparisons). Next, we added gender, its interaction with reading material and order successively. Neither gender nor any of the interactions were significant, and model fit as assessed by the BIC was worse compared to the model without gender, so we did not include gender in later models. As expected, including the ART improved model fit (BIC), $\chi^2_{\text{change}} = 6.0$, $p = .014$. Covariance parameters and standard errors are also smaller (see Table 3), indicating that the model accounts for more variance in RME scores.

Next we added the AQ-S as a fixed covariate. Addition of the AQ-S lowered the variance and standard error of the RME after fiction. However, it was not a significant predictor of RME, and the model fit did not change, so we excluded it from the subsequent analysis. Finally, we added the TS to account for the extent to which participants were engaged with the reading material. In the final model, both the interaction (order*reading material) and the effect of reading material given order were statistically significant. When order, familiarity with fiction, autism quotient scores, and transportation into the narrative were controlled, reading fiction resulted in higher scores on the RME than did reading nonfiction, $d = 0.22$.

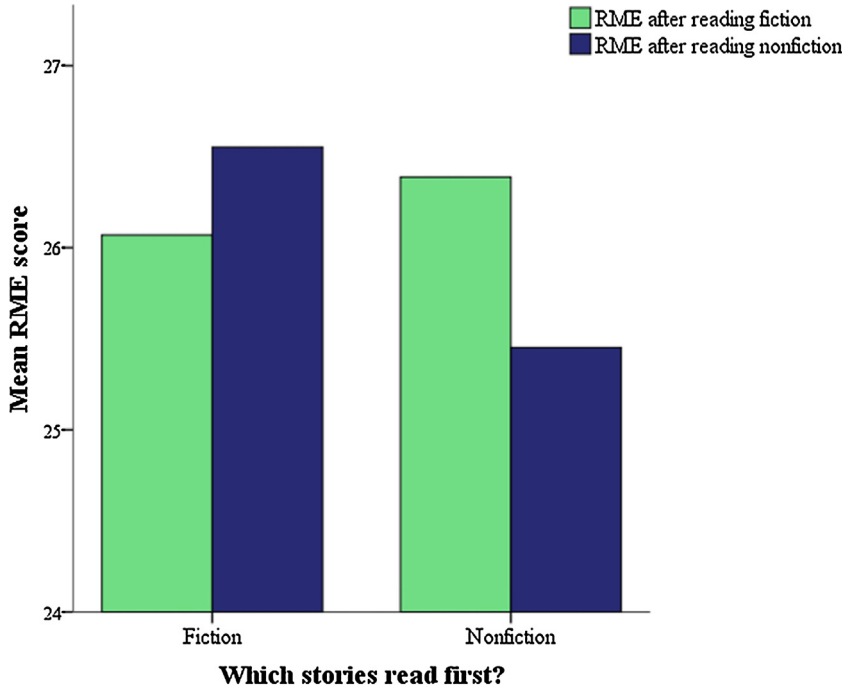

Fig. 1. Initial analyses revealed possible order effects. Participants did better the second time they took the RME.

2.3. Intuitive physics (IP)

We next tested the effect of measured variables and the manipulation on IP scores. Models were tested in the same fashion as described above for the RME; in this case, however, we failed to find an effect. Even controlling for all other variables, type of reading material did not predict scores on the IP, $p = .880$. The strongest predictor of IP scores was gender, with men outperforming women in all models. When only reading material, order, and gender were entered in the model, there was an interaction between gender and order, $F(56) = 4.96$, $p = .030$. Men tended to do better on the IP after reading fiction independently of whether they had already taken the test, whereas women tended to do better the second time they took the test only if they had read first fiction, then nonfiction. However, once covariates were added to the model, the

Table 3
Model comparisons for the effect of type of reading material ("read"), order, gender, and covariates ART, AQ-S, and TS on RME scores. Parameter estimates and corresponding standard errors are for RME scores after reading fiction and nonfiction.

Model	Residual covariance parameter estimate (Standard error)			Fit		*p* value for order*type interaction	Effect of read given order		
	Fiction	Nonfiction	covariance	−2 log Likelihood	BIC		*p*, one-tailed	*t*	*d*
Order, read, order*read	17.55 (3.20)	19.32 (3.53)	13.77 (2.97) [7.96, 19.59]	641	670	.077	0.046	1.71	0.16
Order, read, order*read, gender	17.53 (3.20)	19.34 (3.53)	13.77 (2.97) [7.95, 19.59]	641	674	.077	0.046	1.71	0.16
Order*read*gender	16.62 (3.03)	18.85 (3.44)	13.15 (2.85) [7.60, 18.73]	637	682	.300	0.147	1.06	0.10
Order, read, order*read, ART	14.72 (2.69)	17.99 (3.30)	11.70 (2.59) [6.62, 16.77]	632	**664**	.077	0.046	1.71	0.17
Order, read, order*read, ART, AQ-S	14.30 (2.61)	**17.66 (3.24)**	**11.32 (2.52)** [6.38, 16.26]	630	667	.077	0.046	1.71	0.17
Order, read, order*read, ART, TS	**14.25 (2.61)**	18.77 (3.50)	12.08 (2.64) [6.90, 17.25]	**629**	665	.035	0.011	2.33	0.22

Note. $N = 60$. Best fit and lowest covariance/standard error estimates in bold. 95% Confidence Intervals for covariance estimates are presented in brackets. ART: Author Recognition Task (Acheson et al., 2008); AQ-S: Short Autism Quotient (Allison et al., 2012); TS: Transportation Scale (Green & Brock, 2000).

interaction was no longer significant. Because the ART (familiarity with fiction) had the strongest zero-order correlation with IP scores (see Table 2), it was the first covariate entered. When added to the model along with reading material, gender, order, and the gender*order interaction, it was a significant predictor of IP scores, $F(55) = 6.40$, $p = .014$. Neither transportation nor scores on the AQ-S were significant predictors when added to the model, and they changed neither the fit nor the effect of reading material.

3. Discussion

This experiment had two goals: to replicate Kidd and Castano (2013)'s result concerning the effect of reading literary fiction versus nonfiction on ToM, and to further explore whether reading literary fiction versus nonfiction had an effect—either positive or negative—on performance on a test focused on physical reasoning. Controlling for order, familiarity with fiction, transportation into narratives, and autism quotient scores, we found that participants performed better on the RME after reading fiction than they did after reading nonfiction. In contrast, controlling for these variables, we found no effect of the kind of material a person had read (literary fiction versus nonfiction) on performance on the Intuitive Physics Test. Thus, it does not seem that the increased performance on the ToM task comes at a cost to intuitive physics reasoning, as one might predict based on theories that suggest that the ability to interpret the emotions and mental states of others may compete in some way with the ability to understand physical, rules-based systems (Baron-Cohen, 2009; Manson & Winterbottom, 2012; Wakabayashi et al., 2007; Wright & Skagerberg, 2012). Nor does it appear that reading fiction—even challenging, literary fiction, which presents a sequence of character-centric events that can be seen as complex, ambiguous, and requiring more effort on behalf of the reader (e.g. Koopman & Hakemulder, 2015)—improves performance on cognitive tasks across the board.

These results are noteworthy for several reasons. First, this paper serves as a replication of Kidd and Castano (2013), wherein participants who read literary fiction scored higher on the Reading the Eyes in the Minds Test than participants who read nonfiction. It is worth noting that the effect size in the current experiment was small, even after controlling for other variables. Kidd and Castano (Study 1) reported a larger effect size in a between-subjects design. This may be due in large part to the presence of order effects in the current experiment: participants did better the second time they took the RME, and this improvement was more pronounced if they had read literary fiction first and nonfiction second than the reverse. This pattern of results may be indicative of carryover effects, wherein the facilitation effect of reading two literary short stories in the first block carries over the to second, combining with a general practice effect on the RME to lead to a greater increase of scores. Future research is needed to investigate the duration of these effects and examine whether reading literary fiction, in addition to increasing performance on emotion reading tasks, can enhance learning on these tasks as well.

A second contribution of this work is that it represents the first study that we know of to examine the effects of reading literary fiction (versus nonfiction) on both social cognition *and* physical reasoning. These results are in line with dissertation work by Mar (2007), where a similar pattern of results was found when comparing the effect of reading fiction (versus nonfiction) on social versus analytical reasoning. In this experiment, analytical reasoning was operationalized using questions from the LSAT, which tests deductive reasoning ability and logic. As we did, Mar found that participants performed better on the social reasoning task after reading fiction than after reading nonfiction, but that there was no effect on nonsocial reasoning.

Although it has been suggested that reading serves as social simulation (e.g., Mar & Oatley, 2008; Mar et al., 2006), other scholars have suggested that fiction can also be used to communicate non-social information as well. For example, Sugiyama (1996) argued that stories affect the audience's representations not only of social stimuli, but also of the physical environment, and other scholars have noted the importance of causality in storytelling (e.g., Losh & Capps, 2003). From this perspective, it is interesting that reading fiction did not affect performance on the Intuitive Physics test, which taps participants' ability to reason about physical causality. One explanation for these results requires distinguishing between the *content* we learn from stories and the *skills* that engaging with fictional narratives—particularly literary fiction—may allow us to practice. It is possible that readers may extract information about the physical world from fictional narratives without these narratives priming, simulating, or honing problem-solving and reasoning in the physical domain more generally. A second possibility is that fiction *could* serve to simulate and hone physical reasoning abilities if a narrative focused on a protagonist reasoning his or her way through a physical dilemma, but that the content of fiction—particularly complex literary fiction—may be overwhelmingly more likely to focus on complex social situations, rather than the physical equivalent.

As always, a null result such as this one should be interpreted with caution. In the current experiment, IP scores were most strongly correlated with participant gender: males performed better than females. Because many participants find the Intuitive Physics test to be extremely challenging, it could be that performance on this task is particularly difficult to move around. Although prior research has shown a variety of circumstances under which performance on the Reading the Mind in the Eyes Test is affected by experimental manipulations, such as when a participant is given intranasal oxytocin (Domes, Heinrichs, Michel, Berger, & Herpertz, 2007; Guastella et al., 2010), less work has investigated the circumstances under which performance on the Intuitive Physics Test can be facilitated or impeded by experimental manipulations.

One possible interpretation of the prior work on the relationship between reading fiction and ToM is that reading fiction—particularly challenging literary fiction—may put the reader into "people mode" and prime them to pay attention to the nuances that distinguish between different mental states and emotions. Under this view, it might be expected that narratives

that focus on people, but happen to be true, such as memoirs, may yield similar effects (Koopman & Hakemulder, 2015). From this perspective, it is interesting that reading a nonfiction article that focused on non-social content did not improve IP scores. It is worth noting, however, that the nonfiction passages participants read in this experiment were both about animals. Although animals are not people, they also are not physical objects, but rather, animate creatures. It is possible that if participants had read nonfiction that focused more strongly on objects and the physical mechanics thereof, a different pattern of results may have been found.

Although we found no experimental effect of reading material on performance on the Intuitive Physics Test, our correlational results suggest that the relationship between lifelong reading habits and ToM ability (Kidd & Castano, 2013; Fong et al., 2013; Mar et al., 2006; Mar, Oatley, et al., 2009; Mar, Tackett, et al., 2009) may extend to non-social cognitive ability. Scores on the ART were the strongest predictor of RME performance in our study, as had been the case for Kidd and Castano ($\omega_p^2 = .13$ in our study and theirs; for IP, the effect was $\omega_p^2 = .12$). Notably, familiarity with fiction (scores on the ART) also had a moderate zero-order correlation with IP scores ($r = .30, .35$; see Table 2), and was the strongest predictor in our mixed model after gender. In fact, the relationship was similar to that of familiarity with fiction and ToM ability.

Thus, although the one-time exposure to literary fiction (versus nonfiction) affected only ToM scores, a habit of reading fiction seems to be related to both ToM and intuitive physics understanding. One possibility is that another variable may underlie the relationship between ART scores and performance on the IP. For example, general intelligence, may contribute to both increased literary exposure and causal reasoning ability. A second variable that may underlie the relationship between ART scores and performance on the IP is lifetime exposure to nonfiction. Although the ART has been used by prior fiction research as a measure of lifetime *fiction* exposure (e.g., Kidd & Castano, 2013), ART scores may be related to print exposure more broadly. The ART was designed by creating a list of authors who are primarily known for fictional books (Acheson et al., 2008), but some authors on the list have also written non-fiction, and indeed, Acheson and colleagues found that while ART scores were *most* predictive of self-reported time spent reading fiction, they also predicted time spent reading nonfiction to a lesser degree. As such, it may be that the ART is capturing lifelong exposure to nonfiction, and which in turn may be related to performance on the Intuitive Physics Test.

Another intriguing possibility is that lifetime exposure to fiction per se may be related to the ability to reason about cause and effect. It could either be the case that people who are already at home with reasoning about ruled-based systems may be drawn to fictional stories, particularly those that operate based on more predictable rules than reality, or that reading certain types of fiction could enhance causal reasoning. Causality is a vital element of story-telling (Briner, Virtue, & Kurby, 2012), and it is often left to the reader to interpret the causal relationships from the characters' motivations and intentions, from spatial and temporal clues, and according to expected conventions of narrative in general and of genre in particular (e.g., Bruner, 1991; Zwaan, Langston, & Graesser, 1995). From this perspective, it is worth noting that while our experimental manipulation focused on the effect of reading *literary* fiction, the ART includes authors who write in a variety of genres, including not only literary fiction, but also romance, science fiction and fantasy, and mystery/thriller. It is possible that exposure to literary fiction does not affect the ability to reason about rules-based systems, but that exposure to popular genre fiction *does*.

Genre fiction—popular or mass market—is typified by certain patterns that follow established rules: people read a given genre expecting a certain type of plot, certain tropes and character archetypes, and a specific style of writing (see Carroll, 1994). In a mystery novel, the most obvious suspect is rarely the correct one; in romance novels, the hero and heroine will come together, experience a crisis that threatens their relationship, and typically overcome that crisis to reunite in the end. Thus, it is possible that experience with the well-defined "systems" of genre fiction may either enhance or appeal to the ability to reason about rules-based systems, while literary fiction may, rather than appealing to familiar tropes and rules, engender a level of distance and detachment from the text that allows for deeper contemplation of the social content it contains (see Koopman & Hakemulder, 2015). Future research is needed to examine whether the relationship between ART scores and IP scores is due to a third variable, such as intelligence or exposure to nonfiction, or whether it is being driven by experience with certain *types* of fiction.

Another result that merits further investigation is the fact that, in the current experiment, the negative correlation between Autism Quotient scores and performance on the Reading the Eyes in the Minds test was weaker after reading literary fiction than after reading nonfiction. Although these results are preliminary, they raise the question of whether reading literary fiction could interfere with the well-established correlation between autism deficits and ToM (e.g., Baron-Cohen, 2010). Prior research has shown that individuals on the autism spectrum show a significant preference for nonfiction about objects (Barnes, 2012); there is need for future research that directly examines the effects, if any, of reading literary fiction on ToM in high-functioning individuals on or near the autism spectrum.

4. Limitations and suggestions for future research

Although our results support past experimental (Kidd & Castano, 2013) and correlational (Fong et al., 2013; Mar et al., 2006; Mar, Oatley, et al., 2009; Mar, Tackett, et al., 2009) research, the current study has a variety of limitations, many of which can be attributed to the use of a within-subject design. Many participants were unwilling or unable to read four separate pieces over a three-hour period, even with breaks, and we had to discard one third of the original sample. As such, our results may generalize only to people who are very conscientious or who really enjoy reading. It is important to discover if reading literary fiction affects *all* readers in a similar fashion, or if the social-cognitive benefits are limited to a subset of the

general population. Those participants willing or capable of completing the entire study showed signs of practice effects on the RME, further limiting the conclusions we could draw. Entering order into the model allowed us to test the hypotheses, but further study is needed. Future research on these questions implementing a within-subject design may benefit from spacing apart the individual tests and more effectively incentivizing participants to thoroughly read the passages they are given.

Additional, a methodological innovation of the current research related to our within-in subjects design merits discussion. One important difference between the current experiment and the experiment by Kidd and Castano (2013) is that in the current experiment, participants were told prior to reading that they would be expected to provide a summary of what they had read, and after reading, they were expected to do just that. Although this aspect of our design—which was implemented to discourage participants from skimming as the experiment progressed—did not interfere with the established relationship between reading literary fiction and RME scores, it is worth considering that participants across experiments may read in different ways depending the instructions they are given and that reading in a laboratory setting and reading for pleasure may vary more broadly. In particular, participants who are required to read for class or expected to provide summaries of what they have read may be less likely to read esthetically (see Rosenblatt, 1978) and more likely to focus on extracting information. Future research is needed to investigate not only the effect of *what* participants read, but also *how* they read on ToM ability.

Another limitation of the current work involves the demographics of the participants tested. Our sample was relatively small and largely female. Small samples may bias covariance estimates and result in inflated correlations (Schönbrodt & Perugini, 2013), and the preponderance of female participants limits our ability to draw conclusions about gender differences. Nonetheless, our results hinted at intriguing gender differences with intuitive physics: men tended toward doing better on the IP after reading fiction than after reading nonfiction, regardless of order of presentation. In contrast, women tended to do better the *second* time they took the IP, but only if they had read fiction first followed by nonfiction. Although the pairwise comparisons were not significant, the omnibus interaction test, together with the surprising trends, suggest that a larger, more balanced sample allowing more in-depth investigation of potential gender differences could reveal interesting effects.

Future research is also needed to examine the effect of reading literary fiction (versus nonfiction) on different dependent variables. A great deal of prior research on the relationship between fiction and ToM has used the RME as its dependent measure, and we chose this measure because we were interested in replicating and extending prior research. However, there is a need for research that investigates other aspects of social cognition, such as reasoning about complex moral dilemmas or detecting deception, as well as research that measures not only the ability to correctly identify mental states and emotions, but also the ability to function in social situations and respond appropriately to the emotions. Similarly, although we chose the Intuitive Physics Test because of prior research that has used this test, alongside the RME, to investigate individual differences in neurotypical populations, performance on this one test cannot be taken as representative of the entire domain of non-social cognition. Further, there are cognitive abilities that do not focus on either reasoning about people or reasoning about objects, such as memory, which also merit further exploration.

Finally, certain limitations should also be considered with respect to our narrative stimuli. First, the literary fiction and nonfiction texts used in this experiment varied not only on the basis of fictionality, but also in content, narrativity, and literariness. Second, the nonfiction articles chosen for this experiment focused on animals, rather than physical systems. And finally, this experiment focused on the effects of reading *literary* fiction and did not explore the effects of reading *popular* fiction. As mentioned above, it may be that different genres have different effects on cognitive abilities, both long term and immediate. Testing different types of fiction (see Kidd & Castano, 2013) and nonfiction, as well as different media, such as film and persuasive essay, could add to our understanding of the potential effects of reading material.

5. Concluding remarks

This study tested the effect of reading literary fiction vs. nonfiction on cognitive abilities by assessing the extent to which reading material affects ToM and performance on a task involving physical reasoning in a repeated measures experimental framework. Controlling for other variables, reading literary fiction increased ToM but not physical, rules-based reasoning; however, the strongest predictor of performance on both the RME and the IP was familiarity with fiction. These results may have applications for education and the role that literature can play in facilitating cognitive engagement across domains.

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*, 278–289 DOI: 10.3758BRM.40.1.278.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1000 cases and 3000 controls,. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*, 202–212.

Aristotle (350 BCE) (1902). *Poetics. (S. H. Butcher, Trans)*. London: MacMillan Retrieved from: http://classics.mit.edu/Aristotle/poetics.html

Barnes, J. L. (2012). Fiction, imagination, and social cognition: Insights from autism. *Poetics, 40*, 299–316.

Baron-Cohen, S. (2009). Autism: The empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Science, 1156*, 68–80. http://dx.doi.org/10.1111/j.1749-6632.2009.04467.x

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. In I. Savic (Ed.), *Sex differences in the human brain, their underpinnings and implications* (pp. 167–176). Amsterdam: Elsevier.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society, Series B, Special issue on "Autism: Mind and Brain, 358*, 361–374. http://dx.doi.org/10.1098/rstb.2002.1206
Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163–175.
Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology Psychiatry, 42*, 241–251. http://dx.doi.org/10.1111/1469-7610.00715
Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17.
Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Developmental and Learning Disorders, 5*, 47–78.
Briner, S. W., Virtue, S., & Kurby, C. A. (2012). Processing causality in narrative events: Temporal order matters. *Discourse Processes, 49*, 61–77. http://dx.doi.org/10.1080/0163853X.2011.607952
Brosnan, M. J., Gwilliam, L. R., & Walker, I. (2012). Brief report: The relationship between visual acuity, the embedded figures test and systemizing in autism spectrum disorders. *Autism and Developmental Disorders, 42*, 2491–2497. http://dx.doi.org/10.1007/s10803-012-1505-0
Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18*, 1–21.
Carroll, N. (1994). The paradox of junk fiction. *Philosophy and Literature, 18*, 225–241. http://dx.doi.org/10.1353/phl.1994.0054
Clark, C., & Rumbold, K. (2006). *Reading for pleasure: A research overview*. London: National Literacy Trust.
Demidenko, E. (2004). *Mixed models: Theory and applications*. USA: Wiley-Interscience.
Djikic, M., Oatley, K., & Moldoveanu, M. (2013). Reading other minds: Effects of literature on empathy. *Scientific Study of Literature, 3*, 28–47. http://dx.doi.org/10.1075/ssol.3.1.06dji
Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biological Psychiatry, 61*, 731–733. http://dx.doi.org/10.1016/j.biopsych.2006.07.015
Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170–177.
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), 370–376.
Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Personality and Social Psychology, 79*, 701–721.
Goldenfeld, N., Baron-Cohen, S., & Wheelwright, S. (2005). Empathizing and systemizing in males, females, and autism. *Clinical Neuropsychiatry, 2*, 338–345.
Guastella, A. J., Einfeld, S. L., Gray, K. M., Rinehart, N. J., Tonge, B. J., Lambert, T. J., & Hickie, I. B. (2010). Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biological Psychiatry, 67*, 692–944. http://dx.doi.org/10.1016/j.biopsych.2009.09.020
Howell, D. C. (2010). *Mixed models for repeated (longitudinal) data*. Retrieved from http://www.uvm.edu/~dhowell/methods7/Supplements/Mixed%20Models%20Repeated/Mixed%20Models%20for%20Repeated%20Measures1.pdf
Kidd, D. C., & Castano, E. (2013, October). Reading literary fiction improves theory of mind. *Science*. http://dx.doi.org/10.1126/science.1239918
Koopman, E. M., & Hakemulder, F. (2015). Effects of literature on empathy and self-reflection: A theoretical-empirical framework. *Literary Theory, 9*, 79–111.
Krajmer, P., Janosikova, D., Špajdel, M., & Ostatnikova, D. (2010). Empathizing, systemizing, intuitive physics and folk psychology in boys with Asperger syndrome. *Activitas Nervosa Superior Rediviva, 52*, 57–61.
Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine, 34*, 911–924. http://dx.doi.org/10.1017/S0033291703001624
Losh, M., & Capps, L. (2003). Narrative abilities in high-functioning children with autism or Asperger's Syndrome. *Journal of Autism and Developmental Disorders, 33*, 239–251.
Malle, B. F. (2006). Folk theory of mind: Conceptual foundations of social cognition. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 225–255). Oxford, United Kingdom: Oxford University Press http://dx.doi.org/10.1093/acprof:oso/9780195307696.003.0010
Manson, C., & Winterbottom, M. (2012). Examining the association between empathising, systemising, degree subject and gender. *Educational Studies, 38*, 73–88. http://dx.doi.org/10.1080/03055698.2011.567032
Mar, R. A. (2007). *Simulation-based theories of narrative comprehension: Evidence and implications* Unpublished doctoral dissertation University of Toronto, Toronto, Ontario, Canada.
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. http://dx.doi.org/10.1111/j.1745-6924.2008.00073.x
Mar, R. A., Oatley, K., Hirsh, J., de la Paz, J., & Peterson, J. (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*, 694–712. http://dx.doi.org/10.1016/j.jrp.2005.08.002
Mar, R. A., Oatley, K., & Peterson, J. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications, 34*, 407–428. http://dx.doi.org/10.1515/COMM.2009.025
Mar, R. A., Tackett, J. L., & Moore, C. (2009). Exposure to media and theory-of-mind development in preschoolers. *Cognitive Development, 25*, 69–78. http://dx.doi.org/10.1016/j.cogdev.2009.11.002
Quene, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*, 103–121.
Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. Carbondale, IL: Southern Illinois University Press.
Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*, 609–612.
Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464. http://dx.doi.org/10.1214/aos/1176344136
Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Educational and Behavioral Statistics, 24*, 323–355.
Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*, 402–433 Retrieved from: http://www.rfwest.net/Site_2/Welcome_files/Stanovich-expos-print-RRQ89.pdf
Sugiyama, M. S. (1996). On the origins of narrative: Storyteller bias as a fitness-enhancing strategy. *Human Nature, 7*, 403–425. http://dx.doi.org/10.1007/BF02732901
Wakabayashi, A., Baron-Cohen, S., Uchiyama, T., Yoshida, Y., Kuroda, M., & Wheelwright, S. (2007). Empathizing and systemizing in adults with and without Autism Spectrum Conditions: Cross-cultural stability. *Journal of Autism and Developmental Disorders, 37*, 1823–1832. http://dx.doi.org/10.1007/s10803-006-0316-6
Wakabayashi, A., Sasaki, J., & Ogawa, Y. (2012). Sex differences in two fundamental cognitive domains: Empathizing and systemizing in children and adults. *Individual Differences, 33*, 24–34.
Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., Weil, L., & Wakabayashi, A. (2006). Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-R) and empathy quotient (EQ). *Brain Research, 1079*, 47–56. http://dx.doi.org/10.1016/j.brainres.2006.01.012
Wright, D. B., & Skagerberg, E. M. (2012). Measuring empathizing and systemizing with a large US sample. *PLoS ONE, 7*, e31661. http://dx.doi.org/10.1371/journal.pone.0031661
Zunshine, L. (2003). Theory of mind and experimental representations of fictional consciousness. *Narrative, 11*, 270–291.

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

Stimuli used

Delillo, D. (1988, September). The runner. *Harper's Magazine,* 61–63.
Dunn, R. (2012). The story of the most common bird in the world. *Smithsonian Magazine*, *2*(March).. Retrieved from: http://www.smithsonianmag.com/
Saunders, G. (2007, May). *Puppy*. The New Yorker Retrieved from: http://www.newyorker.com/
Vanderbilt, T. (2013, October). *The CIA's most highly-trained spies weren't even human*. Smithsonian Retrieved from: http://www.smithsonianmag.com/

Jessica Black earned a Master's degree in Research Psychology at Humboldt State University. She is now a doctoral student at University of Oklahoma, studying issues at the intersection of narrative and morality, under Dr. Jennifer Barnes. Her interests include moral cognition, philosophy, and quantitative psychology.

Jennifer Barnes is an Assistant Professor of psychology at the University of Oklahoma. In addition to researching the psychology of fiction and the imagination, she is also a professional novelist and television writer. She received her Ph.D. from Yale University in 2012. She is a member of the Writers Guild of America and a Fulbright Scholar.