

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21 On Literary Fiction and its Effects on Theory of Mind

22 David Kidd, Martino Ongis, and Emanuele Castano

23
24
25
26 The New School for Social Research

27
28
29
30
31
32
33 Author Note

34 David Kidd, Department of Psychology, The New School for Social Research

35
36 Martino Ongis, Department of Psychology, The New School for Social Research

37
38 Emanuele Castano, Department of Psychology, The New School for Social

39
40
41
42 Research

43
44 This research was supported by the Whiting Foundation. The authors thank Giulia
45 Baggio for helping edit and prepare the manuscript for publication.

46
47
48 Correspondence should be addressed to David Kidd, Martino Ongis, or Emanuele
49
50 Castano, Department of Psychology, New School for Social Research, 80 Fifth Avenue,
51
52 New York, NY 10011. Contact: kidd305@newschool.edu, ongim671@newschool.edu,
53
54
55
56
57
58 castano@newschool.edu

Abstract

Storytelling is a hallmark human activity. We use stories to make sense of the world, to explain it to our children, to create communities, and to learn about others. This article focuses on fictional stories and their impact on complex sociocognitive abilities. Correlational and experimental evidence shows that exposure to fiction recruits and hones our ability to represent others' mental states, or theory of mind (ToM). Experimental studies suggest this effect is specific to literary fiction. Using a unique set of texts, we replicate the finding that literary fiction improves ToM performance. Consistent with the expectation of greater focus on characters in literary fiction, linguistic analysis of the texts revealed that the literary texts contain more markers of reflective function, a sophisticated manifestation of ToM. Further analysis showed the prevalence of markers of reflective function partially mediated the effect of literary fiction on ToM performance.

Keywords: literary fiction, genre fiction, theory of mind, reflective function, text analysis, interpersonal perception

On Literary Fiction and Its Effects on Theory of Mind

Humans started becoming human about one million years ago. Around this time, the complex social behaviors and institutions that characterize our species began to emerge alongside the development of sophisticated cognitive abilities, including theory of mind (ToM) (Wiessner, 2014). About a million years ago is also when humans began controlling fire. Aside from other important benefits (e.g. medium-rare steak, s'mores), early humans' control of fire enabled them to artificially extend the day. How did they use this extra time? According to anthropologist Polly Wiessner, they did not simply do more of what they did during daylight. Rather, they began telling stories. "[F]irelit activities centered on conversations that evoked the imagination [and] helped people remember and understand others in their external networks" (Wiessner, 2014; p. 1). Fires were brought inside the house, and humans eventually gathered around flashing screens instead of fires: large ones at first and very tiny ones in more recent times. Yet the appetite for stories stayed with us, and is now fed by the very cultural artifacts that our ability to tell stories created a need for and brought about. Among them are plays, movies, and books.

Scholars in the humanities and social sciences have long investigated the functions that these cultural products serve (Humphrey, 1984). Here we focus on one of these cultural artifacts, print fiction, and its effects on the psychological capacity to explain and predict others' behavior on the basis of their intentions, knowledge, and emotions. This ability to infer and understand others' mental states, known as Theory of Mind (ToM; Heyes & Frith, 2014), facilitates the complex cooperative relationships that typify our social lives (Dunbar, 2003). In particular, ToM plays a critical role in perspective-taking, a cognitive component of empathy, and affective empathy when the others' feelings are not immediately apparent, are unfamiliar, or are complex

(Mitchell & Phillips, 2015; Singer, 2006). Drawing from literary theory and extant psychological research on fiction, we propose that reading fiction can be an exercise in advanced ToM that prompts readers to represent and engage with characters' nuanced mental states. We further argue for the need to differentiate between types of fiction, present data in support of the hypothesis that it is specifically literary fiction that fosters ToM, and provide initial empirical evidence of the features of literary fiction responsible for this effect.

Fiction and Mentalizing

Over the past decade, psychologists have begun testing the theory that readers understand the characters and relationships in fiction using the same psychological processes they use when navigating the real social world. Mar and Oatley (2008) portray reading fiction as a means of simulating social experience, with chronic readers gaining social knowledge and honing key sociocognitive abilities. Correlational research supports this view: Individuals who read more fiction perform better on tests of ToM (Mar, Oatley, Hirsh, de la Paz, & Peterson, 2006; Mar, Oatley, & Peterson, 2009). Neuropsychological evidence further bolsters the simulation account, revealing greater activity in mentalizing networks when participants read literary depictions of mental states than when they read physical descriptions, and linking the degree of mentalizing network activation to performance on a test of ToM (Tamir, Bricker, Dodell-Feder, & Mitchell, 2016).

We agree with the social simulation account of fiction, and further propose that different sorts of fiction make varying demands on distinct sociocognitive abilities as readers simulate their fictional social worlds. Specifically, we argue that not all fiction is equally likely to evoke ToM. Psychologists have shown that social contexts do not inevitably activate ToM processes (for a review, see Apperly, 2012). Indeed, the default manner of social perception does not

1
2
3
4 reliably involve understanding others as individuals, with their own unique thoughts, feelings
5
6 and motives. Instead, individuals commonly navigate their social worlds by drawing on more
7
8 efficient processes such as stereotyping and the application of social roles and scripts
9
10 (Hirschfeld, 2001). Social psychologists refer to perception guided by attention to what makes
11
12 others unique as *individuation* and perception based on social identities and roles as
13
14 *categorization* (Brewer, 1988; Fiske, Lin, & Neuberg, 1999; Fiske & Neuberg, 1990; Macrae &
15
16 Bodenhausen, 2000; Young, Hugenberg, Bernstein, & Sacco, 2012). Individuation is likely to
17
18 occur only when the perceiver is especially motivated (e.g., in cases of outcome dependency;
19
20 Neuberg & Fiske, 1987) or when their attention is directed toward the subjective states of the
21
22 target (Galinsky & Moskowitz, 2000; Wheeler & Fiske, 2005). Otherwise, category-based social
23
24 perception generally occurs. We posit that the simulated social worlds of fiction are likely to
25
26 evoke ToM processes when they render category-based social perception unreliable or
27
28 insufficient, notably by disrupting schematic understandings of situations and actors, and by
29
30 drawing attention to characters' complex mental states.
31
32
33
34
35
36
37

38 Popular genre fiction, we submit, often draws upon and reifies readers' knowledge of
39
40 archetypal or stereotypical characters. This may help explain why it is so easily read: It relies on
41
42 largely prefabricated characters that are understood using highly efficient sociocognitive
43
44 processes, freeing up resources for enjoying the development of the plot. Aside from
45
46 appreciating the plot, a reader of a work of popular genre fiction may also derive satisfaction
47
48 from the accuracy of their judgments about the relatively simple and stable characters (Gerrig &
49
50 Rapp, 2004)—a rare pleasure in real life. For example, the Jack Ryan character in Tom Clancy's
51
52 novels is clean-cut, intelligent, physically fit, and unflinchingly patriotic. The stability and
53
54 cohesiveness of these characteristics makes it easy to follow his thrilling adventures without
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 stopping to ask nuanced questions about his motives. In contemporary literary fiction, by
5
6 contrast, plot is often secondary, primarily serving to develop characters and reveal their
7
8 psychological complexity. For example, Don DeLillo's depiction of Lee Harvey Oswald's path
9
10 toward assassinating President Kennedy in *Libra* (1988) is focused less on the mechanics of *how*
11
12 Oswald succeeded in murder and more on exploring (without clear resolution) *why* he was
13
14 motivated to do so.
15
16
17

18
19 Popular genre and literary fiction are, of course, contested, partially overlapping
20
21 categories, and what we propose above is to be understood as a depiction of broad and relative
22
23 differences between the two, rather than a clear-cut distinction. Yet, the relative greater emphasis
24
25 on the inner lives of characters, rather than plot development, in literary fiction is what leads us
26
27 to expect reading literary fiction to evoke ToM processes. Initial evidence supports this idea.
28
29 Kidd and Castano (2013) found that participants who were asked to read works of literary fiction
30
31 scored higher than those assigned works of genre fiction on the Reading the Mind in the Eyes
32
33 Test (RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), an advanced test of ToM
34
35 that requires integration of affective and cognitive ToM processes (Mitchell & Phillips, 2015).
36
37 Black and Barnes (2015a) extended these findings using the RMET in a within-participants
38
39 design. While the literary fiction group exhibited improvement, the non-fiction group did not. In
40
41 this study, it was further shown that literary fiction did not boost performance on a cognitively
42
43 demanding task that does not engage *social* perception processes. This evidence aligns with
44
45 theoretical work by cognitive literary theorists, who point to the importance of characters'
46
47 psychology complexity (Culpeper, 2001) and the subtlety of their presentation in literary fiction
48
49 (Zunshine, 2015) to the inferential activity of readers. However, these results are preliminary,
50
51 and no evidence has yet been presented with regard to the characteristics of literary fiction that
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 may be responsible for its effects.
5

6 **What is in a Work of Literary Fiction?**

7
8
9 Although there are many stylistic dimensions along which literary and genre fiction can
10 be compared, here we limit ourselves to a focused analysis of characteristics of the text that are
11 conceptually related to the findings reported above. Specifically, we investigate whether
12 quantifiable linguistic differences between works of literary and genre fiction may help us
13 understand the different effects they have on readers' ToM. Given the theoretical importance of
14 the emphasis on character development relative to plot, we focus on a textual analysis of
15 linguistic indicators of sophisticated explanations of one's own and others' behavior in terms of
16 mental states, or *Reflective Function* (RF; Fonagy & Target, 1997; Slade, 2005). RF is a
17 manifestation of a well-developed capacity for ToM, which is thought to be its "necessary
18 foundation stone" (Knox, 2003, p. 143).
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 The concept of RF is especially relevant to the issue at hand because of its focus on the
34 complexity of the mental models that people generate to understand others (Bouchard et al.,
35 2008). The psychological construction of these models requires individuals to build detailed and
36 nuanced representations of others' perspectives. Unlike more basic tests of ToM, such as false-
37 belief tasks, the measurement of RF differentiates between levels of sophistication of
38 mentalizing, rather than the mere capacity for it. For example, the highest levels of RF involve
39 not only understanding that others have mental states, but an awareness of multiple possible
40 explanations for others' behavior and the limits of one's insight into others' minds (Fonagy,
41 Target, Steele, & Steele, 1998). This construct is thus closely aligned with the sort of
42 interpretation and reflection upon characters that literary fiction is expected to promote. If
43 literary fiction tends to emphasize character development over plot to a greater extent than
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

popular genre fiction, literary texts will be higher in RF than genre texts. Furthermore, we expect that levels of RF will help explain literary fiction's effects on ToM performance.

Empirical Investigations

The research reviewed above provides initial evidence that literary fiction may engage ToM processes. In this contribution, we present (1) an experiment aiming at replicating and extending this finding, and (2) a supplementary analysis beginning to investigate characteristics of literary fiction responsible for it.

Experiment

We begin with a conceptual replication of Kidd and Castano's (2013) findings, but with a larger set of literary and genre texts. Since the hypothesis deals with the broad categories of literary and genre fiction, a set of new texts was added to those used in prior research to further test the generality of the specific effects of literary fiction on ToM.

Participants. Participants ($N = 403$) were recruited online and compensated (\$2) using Amazon.com Mechanical Turk (Mturk), and they completed the experiment on the Qualtrics online platform. Several participants were excluded on the basis of six criteria, the first five of which are the same used by Kidd & Castano (2013): inadequately short reading time for the assigned texts (i.e., an average of < 30 s per full page of text; $n = 40$), unusually long reading time for the text (i.e., > 3.5 SD from mean; $n = 2$), high guessing scores on the measure of familiarity with fiction (i.e., > 3.5 SD from mean; $n = 6$), participation in previous similar studies of ours ($n = 24$), participants for whom English was a second language ($n = 9$), and participants who reported being distracted or having computer problems while completing the experiment ($n = 5$). The final sample was composed of 317 participants (54.89% female), who ranged in age from 19 to 82 ($M = 33.97$, $SD = 11.29$).

Materials and procedure. Participants were randomly assigned to read one of 20 texts (5 literary fiction short stories, 5 literary fiction excerpts, 5 genre fiction short stories, 5 genre fiction excerpts) or to read nothing at all (see Appendix A for a list of all texts). The short stories, with one exception (*Tuesday Night Club* by Agatha Christie) had all been previously used in Kidd and Castano (2013), while the excerpts from novels are original stimuli not previously used in any other research on this topic that we know of.

As in Kidd and Castano (2013), the judgement of expert raters (i.e., literary prize jurors) was used as a criterion for identifying literary fiction. Although other qualities, such as the use of metaphor and other stylistic devices, contribute to literariness (Miall & Kuiken, 1999), character development and complexity is central to contemporary definitions of literariness (Culpeper, 2001; Hakemulder, 2000; Zunshine, 2015). Consequently, we expected winners of literary prizes to focus readers' attention on nuanced characters and their mental states. The literary texts thus included five short stories selected from an anthology of the 2012 PEN/O. Henry Prize winners and excerpts from five contemporary works of fiction that appeared on the longlist for the 2013 National Book Award for fiction.

Contrary to acclaimed works of literary fiction, works of genre fiction focus more on plot than character development. Consequently, characters in such works are more likely to be unidimensional (Forster, 1927) and stereotypical (Culpeper, 2001). Genre texts were thus short stories representing different genres (i.e., romance, science fiction, mystery, and adventure) selected from an edited anthology (Hoppenstand, 1998) and excerpts from recent bestsellers on Amazon.com from the sales categories of romance, mystery (including thriller and suspense fiction), and general genre fiction.

The literary and genre texts did not differ in number of words ($M_{\text{Literary}} = 4504.00$, $SD =$

1523.75, $M_{\text{Genre}} = 5116.00$, $SD = 1536.40$, $F(1,18) = 0.80$, $p = .38$), Flesch Reading Ease ($M_{\text{Literary}} = 80.42$, $SD = 8.32$, $M_{\text{Genre}} = 80.40$, $SD = 5.15$, $F(1,18) = 0.00$, $p = .99$), or Flesch-Kincaid Grade Level ($M_{\text{Literary}} = 5.71$, $SD = 2.31$, $M_{\text{Genre}} = 4.95$, $SD = 0.81$, $F(1,18) = 0.96$, $p = .33$).

Upon finishing the reading task (or immediately upon granting consent, for those in the control condition), participants completed the RMET. This task, which has been used previously in studies of reading and mentalizing (e.g., Black & Barnes, 2015a; Fong, Mullin, & Mar, 2013; Kidd & Castano, 2013; Mar et al., 2006; 2009), includes 36 trials in which an image of an actor's eyes are shown, and the participant must choose which of four complex emotion terms (e.g., sympathetic, irritated, thoughtful, encouraging) best matches the actor's mental state. Participants' scores on this measure are calculated by summing the responses they give that correspond with those given by judges and a large sample of normal adults in the measure's initial validation studies (Baron-Cohen et al., 2001). One of the most widely used and validated measures of advanced ToM, the RMET has been utilized in more than 250 studies, translated into several languages, and adapted for use with children (Baker, Peterson, Pulos, & Kirkland, 2014).

The RMET is considered an advanced measure of ToM because cognitive and affective elements of ToM are necessary to understand the complex mental states it depicts (Baron-Cohen et al., 2001; Mitchell & Phillips, 2015). The advanced ToM abilities required by the RMET likely contribute to its sensitivity to individual differences in mentalizing among neurotypical adults (Baron-Cohen et al., 2001; Vellante et al., 2013), and to experimental manipulations that draw attention to others' mental states, including playing narrative video games (Bormann & Greitemeyer, 2015), watching award-winning TV dramas (Black & Barnes, 2015b), compassion meditation instruction (Mascaro, Rilling, Negi, & Raison, 2012), and reading metaphors (Bowes

1
2
3
4 & Katz, 2015). Research indicates that reading literary passages depicting social content
5
6 activates the same brain regions (Tamir et al., 2016) as the RMET (Schurz & Perner, 2015), such
7
8 as the medial prefrontal cortex and temporoparietal junction. This further suggests the
9
10 appropriateness of the RMET to testing the hypothesized effects of reading on mentalizing.
11
12

13
14 Next, participants completed the Author Recognition Test (ART; Acheson, Wells, &
15
16 MacDonald, 2008). This is a list of 130 names, 65 of which are names of real fiction writers, and
17
18 the task is to select those who are recognized as writers. Scores are calculated by subtracting the
19
20 number of non-authors (foils) selected from the number of authors correctly identified. Previous
21
22 research has shown that ART scores correlate positively with performance on the RMET (Mar et
23
24 al., 2006; 2009; Kidd & Castano, 2013). Finally, participants provided demographic information
25
26 (including age, gender, level of education, first language, and race). They were also asked
27
28 whether they had previously participated in a similar study, and to indicate if they were
29
30 distracted or had computer problems while completing the study.
31
32
33
34

35
36 **Results.** An ANCOVA with condition (literary, genre, baseline), ART scores (centered),
37
38 and their interaction entered as predictors was conducted using RMET as the dependent variable.
39
40 The covariate was ethnicity (White participants [$n = 251$] vs. non-White [$n = 66$]). Previous
41
42 research has demonstrated that participants perform better on the RMET when the stimuli are
43
44 images of in-group rather than out-group members (Adams et al., 2010). Consistent with this
45
46 finding, White participants performed better on the RMET, which depicts only White faces, than
47
48 non-White participants ($F(1,315) = 7.53, p < .01, \eta^2 = .023, 95\% \text{ CI} [.001, .065]$). Thus, we
49
50 included ethnicity as a covariate to reduce within-group error variance (Field, 2009). The
51
52 percentages of White and non-White participants did not differ across conditions ($X^2(2, N = 317)$
53
54 $= 0.09, p > .95$). No other demographic variable (i.e., gender, age, education) was significantly
55
56
57
58
59
60
61
62
63
64
65

related to RMET scores ($ps > .07$). The analysis revealed a significant effect of condition ($F(2,310) = 3.07, p = .04, \eta_p^2 = .019, 95\% \text{ CI} [.000, .055]$). Consistent with prior findings, performance on the RMET was higher in the literary condition (LSMean = 26.22) than in the genre fiction condition (LSMean = 24.71, $t = 2.28, p = .02$) and the baseline condition (LSMean = 24.98, $t = 1.96, p = .05$). Both covariates were significant, with higher scores on the ART predicting better RMET performance ($\beta = .25, p < .01, \eta_p^2 = .067, 95\% \text{ CI} [.022, .124]$), and White participants (LSMean = 24.64) performing better on the RMET than non-White participants (LSMean = 24.64, $F(1,310) = 4.15, p = .04, \eta_p^2 = .013, 95\% \text{ CI} [.000, .048]$).

One aim of this experiment was to extend prior experimental findings by using new texts. In order to directly test whether the new literary texts led to improvements in ToM performance relative to the new popular genre fiction texts, a separate analysis was conducted using only participants assigned to read one of the 10 new excerpts. The analytical strategy was the same as above, and it yielded similar effects. There was a marginal effect of condition ($F(2,198) = 2.65, p = .07, \eta_p^2 = .026, 95\% \text{ CI} [.000, .076]$). Performance on the RMET was significantly higher in the literary condition (LSMean = 26.44) than in the genre condition (LSMean = 24.75, $t = 2.03, p < .05$) and the baseline condition (LSMean = 24.96, $t = 1.93, p = .05$). Again, higher scores on the ART predicted better RMET performance ($\beta = .30, p < .01, \eta_p^2 = .095, 95\% \text{ CI} [.031, .174]$), and White participants (LSMean = 26.54) performed better on the RMET than non-White participants (LSMean = 24.23, $F(1,198) = 9.19, p < .01, \eta_p^2 = .044, 95\% \text{ CI} [.005, .109]$).

Supplementary Analysis

Materials and procedure.

Computerized Reflective Function (CRF). Reflective function is most often measured in relatively small samples of texts by highly trained coders (Fonagy et al., 1998). A Computerized

1
2
3
4 Reflective Function (CRF) method has also been developed (Fertuck, Mergenthaler, Target,
5
6 Levy, & Clarkin, 2012), which provides RF scores for texts based on linguistic markers. To
7
8 develop the CRF, Fertuck et al. (2012) first obtained Reflective Functioning (RF) ratings for
9
10 clinical and non-clinical Adult Attachment Interviews (AAIs) from trained judges using the
11
12 Reflective Functioning Scale (Fonagy et al., 1998). After selecting subsets representing the
13
14 interviews highest and lowest in RF, Fertuck et al. (2012) used the Marker Approach
15
16 (Mergenthaler & Bucci, 1999) to identify words appearing most frequently in low RF interviews
17
18 and those most frequent in high RF interviews. These words were then used to create low and
19
20 high RF dictionaries (see Appendix A). The remaining texts in Fertuck et al.'s sample of AAIs
21
22 were then analyzed using the CRF. In both the clinical and non-clinical samples the RF scores
23
24 generated by the CRF were significantly correlated with those given by the trained raters
25
26 (Fertuck et al., 2012), providing evidence for the criterion validity of the CRF.
27
28
29
30
31
32

33
34 Critically, the method of dictionary creation used in developing the CRF is distinct from
35
36 those based primarily on content validity for selection of words. The RF dictionary contains
37
38 mostly function words that do not relate to mental states, making the RF dictionary different
39
40 from seemingly related dictionaries, such as the social and cognitive dictionaries of the LIWC
41
42 (Pennebaker, Booth, Boyd, & Francis, 2015). This difference is important because people often
43
44 use mental state terms in non-reflective ways. Fonagy et al. (1998) contrast, for example, the
45
46 statement, "***I was** angry, now *I am* sad*", which is not very reflective, with the more reflective
47
48 statement, "***I felt really** angry **and then it** changed *to* sadness*" (p. 20). In the first case, the
49
50 speaker only reports their emotions at different times; but in the second case, the speaker
51
52 demonstrates an awareness of how their emotional response to something transformed over time.
53
54 In the examples above, high RF words in the CRF dictionary are in bold font and low RF words
55
56
57
58
59
60
61
62
63
64
65

are in italics, illustrating how the CRF method identifies indicators of RF. The high RF example yields a CRF score of .30 (High RF (5/10) – Low RF (2/10), while the low RF example yields a score of -.14 (High RF (1/7) – Low RF (2/7). Thus, RF is more a function of how mental states are discussed rather than the mere presence of mental state terms.

In order to test the hypothesis that the degree of RF in the texts correlates with participants' ToM performance, we used the CRF to calculate RF scores for all the new texts used in the current experiment (N = 11) and the texts used in four different experiments by Kidd and Castano (2013; N=12), for which RMET scores are also available. The final sample was composed of 23 texts (13 literary fiction, 10 genre fiction) (see Appendix B). CRF scores were obtained for each text using the DAAP (Maskit, & Murphy, 2014) and the dictionaries identified in Fertuck et al. (2012). For each text a DeltaRF score was obtained by subtracting the LowRF from the HighRF score.

Results. An ANOVA on the DeltaRF score (which was logarithmically transformed to improve normality of the distribution) revealed a significant effect of type of text (literary vs. genre fiction), after removing one text that was both a statistical and a conceptual outlier¹: $F(1, 21) = 3.19, p = .044$, one-tailed. As hypothesized, DeltaRF means² were higher in the literary fiction condition ($M = .20, SD = .05$) than in the genre fiction condition ($M = .17, SD = .05$). A correlation was then computed between DeltaRF and the average RMET scores for each text (obtained by averaging the RMET score of participants who had read each text). This revealed a

¹ The story *Chameleon* by Anton Chekhov was found to have high influence on the results using Cook's distance (i.e., Cook's distance > 4/n). Compared to other literary texts, it had very low RF. This is consistent with Chekhov's avoidance of direct portrayals of mental states in this text. In this particular comedic story, a provincial police superintendent radically changes his attitudes towards a dog depending upon whom he thinks it belongs to, but these shifts are made apparent only in his behavior. The CRF, it seems, is limited in its ability to identify this manner of representing complex mental activity.

² Untransformed means are reported for ease of interpretation.

positive correlation ($r(20) = .48, p = .02$), indicating that the greater the RF of a text, the better its readers performed on the subsequent ToM task.

We further tested this relationship using participant, instead of text, as the unit of analysis. First, RMET scores were regressed on RF values. As expected, there was a positive relation between RF and RMET performance ($\beta = 0.16, t(609) = 4.14, p < .001, R^2 = .027$). We then tested whether RF mediated the effect of type of text (literary vs. genre fiction) on RMET performance using Hayes's (2013) bootstrapping procedure. We observed a significant indirect effect ($b = 0.47, SE = 0.23, 95\% CI [0.05, 0.92], \kappa^2 = .044, 95\% CI [0.006, 0.084]$). The direct effect of condition was also significant ($b = 1.20, SE = 0.43, p < .01, 95\% CI [0.35, 2.04]$), leading us to conclude that RF partially mediated the effect of condition on RMET performance.

Since the CRF is a proxy of the extent to which a text provides sophisticated interpretations of behavior in terms of mental states, these findings are consistent with the argument that literary fiction, relative to genre fiction, tends to be more focused on understanding and developing characters. Interestingly, this difference is not apparent in analyses of the explicit content of literary and popular genre texts. As discussed above, most linguistic analysis tools, such as the LIWC (Pennebaker et al., 2015), assess features of a text by counting words related to specific domains. Most relevant to the present investigation, the LIWC includes measures of the prevalence of words related to social processes (e.g., talk, neighbor) and cognitive processes (e.g., think, maybe). Using LIWC (Pennebaker et al., 2015), we found that the two types of texts did not differ in social and cognitive processes words ($ps > .40$), and these variables did not mediate any of the effect of condition on RMET performance³. Taken together,

³ We thank an anonymous reviewer for suggesting these analyses. We further tested the other general LIWC variables (i.e., function words, affect words, perceptual processes, biological processes, core drives and needs, and relativity), none of which differed across conditions (all $ps > .12$) or mediated the effect of condition.

these results support the expectation that while literary and popular genre fiction equally depict social and cognitive content, they do so in different ways. It is this difference that seems most important to understanding their distinct effects on ToM.

General Discussion

In this article, we reported a replication of the results obtained by Kidd and Castano (2013) showing that reading literary fiction improves ToM performance. Reading genre fiction, on the other hand, did not lead to improvement. Of importance, the results held also when the analyses were conducted on the subsample of texts that had not been used as stimuli in prior research. This result improves the generalizability of previous findings showing that reading literary fiction enhances ToM performance.

Drawing from literary theory, we propose that this occurs because literary texts are more likely to focus on developing characters than the movement of the plot. Using a computerized measure of indicators of RF, a construct closely linked to ToM, we found support for this view. Insofar as this measure is a good proxy of RF, these results suggest that the literary texts include more sophisticated interpretations of behavior in terms of mental states. Moreover, the extent of RF partially mediated the effect of literary fiction on ToM. Together, these results lend support to the underlying theory that readers' attention to the inner lives of characters, rather than plot, improves ToM performance.

This interpretation, though theoretically grounded, must be considered alongside the limitations of the reported research. First, although we used a large number of stimuli relative to many psychological studies, our number of texts is very small compared to the number of fictional works available. Moreover, the difference between the groups of texts on the measure of RF was quite small, and there was substantial variability in RF across texts within the categories

of literary and genre fiction. Future research examining a larger corpus of texts is necessary to better understand whether RF is indeed more characteristic of literary fiction than genre fiction, and a reliable predictor of readers' ToM performance.

Likewise, it must be emphasized that the measure of RF only partially accounted for the effect of reading literary fiction on ToM performance. This may be due to measurement error, as the CRF, though it has criterion validity, is unable to detect important nuances that are important to the RF construct, such as the presence of self-serving biases or inappropriate interpretations. However, it seems more likely that the extent of RF in literary fiction is only one feature among others that is responsible for the effect. Thus, we see these results as a first step towards identifying quantifiable linguistic features in narratives that can recruit ToM processes. Other features of literary fiction, such as metaphor (Bowes & Katz, 2015) and other forms of foregrounding (Miall & Kuiken, 1999), may also contribute to the effect.

It is also crucial to keep in mind that the observed effects are thought to reflect the mental activity of the reader. While we argue that textual features, including the indicators of RF discussed here, help guide readers' engagement with the text, it is also important to further study readers' involvement with texts more directly. For example, participants could be asked to indicate the extent to which they were interested in the plot or character development. Also, individual differences, such as dispositional need for cognition (Cacioppo & Petty, 1982), may make some readers more likely to respond to literary texts than others, as suggested by research indicating greater involvement in texts among readers with high need for cognition (Green et al., 2008). Such research could supplement analyses of texts to present a more nuanced picture of the ways in which readers and texts interact to impact psychological processes of social cognition.

Stories are a hallmark of human culture, and they permeate our daily lives. A growing

1
2
3
4 body of research has begun to shed light on the psychological consequences of our engagement
5
6 with fiction. We propose, along with others (e.g., Fong, Mullin, & Mar, 2015), that research into
7
8 the effects of fiction will be most informative when it accounts for differences, both in content
9
10 and style, among types of stories. In the present research, we have drawn on the view in literary
11
12 theory that a focus on characters is a distinguishing feature of literary fiction. Our findings
13
14 support this view, and further suggest that this feature has an effect on ToM, a key psychological
15
16 process. More broadly, this research points to the importance of adopting an interdisciplinary
17
18 approach to studying the effects of fiction. Such research will help illuminate the importance of
19
20 fictional worlds to our experience of the real one.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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*(1), 278-289.
- Adams Jr, R. B., Rule, N. O., Franklin Jr, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., ... Ambady, N. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97-108.
- Apperly, I. A. (2012). What is “theory of mind”? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology*, *65*(5), 825-839.
- Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and “Reading the Mind in the Eyes”. *Intelligence*, *44*, 78-92.
- 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 and Psychiatry*, *42*(2), 241-251.
- Black, J. E., & Barnes, J. L. (2015a). The effects of reading material on social and non-social cognition. *Poetics*, *52*, 32-43.
- Black, J., & Barnes, J. L. (2015b). Fiction and social cognition: The effect of viewing award-winning television dramas on theory of mind. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(4), 423-429.
- Bormann, D., & Greitemeyer, T. (2015). Immersed in virtual worlds and minds effects of in-game storytelling on immersion, need satisfaction, and affective theory of mind. *Social*

Psychological and Personality Science, 6(6), 646-652.

Bouchard, M. A., Target, M., Lecours, S., Fonagy, P., Tremblay, L. M., Schachter, A., & Stein,

H. (2008). Mentalization in adult attachment narratives: Reflective functioning, mental states, and affect elaboration compared. *Psychoanalytic Psychology*, 25(1), 47.

Bowes, A., & Katz, A. (2015). Metaphor creates intimacy and temporarily enhances theory of mind. *Memory & Cognition*, 43(6), 953-963.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S.

Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116.

Culpeper, J. (2001). *Language and characterisation: People in plays and other texts*. London:

Longman.

Dunbar, R. I. (2003). The social brain: Mind, language, and society in evolutionary perspective.

Annual Review of Anthropology, 32, 163-181.

Fertuck, E. A., Mergenthaler, E., Target, M., Levy, K. N., & Clarkin, J. F. (2012). Development and criterion validity of a computerized text analysis measure of reflective functioning.

Psychotherapy Research, 22(3), 298-305.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S.

Chaiken & Y. Trope (Eds.), *Dual-process Theories in Social Psychology* (pp. 231-254). New York, NY: Guilford.

Fiske, S. T., & Neuberg, S. E. (1990). A continuum of impression formation, from category-

based to individuating processes: Influences of information and motivation on attention

and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1_74). San Diego, CA: Academic Press.

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.

Fong, K., Mullin, J. B., & Mar, R. A. (2015). How exposure to literary genres relates to attitudes toward gender roles and sexual behavior. *Psychology of Aesthetics, Creativity, and the Arts*, 9(3), 274-285.

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology*, 9(4), 679-700.

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual, version 5.0, for application to Adult Attachment Interviews*. London, UK: University College London.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, 78(4), 708-724.

Gerrig, R. J., & Rapp, D. N. (2004). Psychological processes underlying literary impact. *Poetics Today*, 25(2), 265-281.

Green, M. C., Kass, S., Carrey, J., Herzig, B., Feeney, R., & Sabini, J. (2008). Transportation across media: Repeated exposure to print and film. *Media Psychology*, 11(4), 512-539.

Hakemulder, J. (2000). *The moral laboratory: Experiments examining the effects of reading literature on social perception and moral self-concept*. Philadelphia: John Benjamins.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A*

1
2
3
4 *regression-based approach*. New York, NY: Guilford Press.

5
6
7 Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190),
8
9 1243091-1–1243091-6.

10
11 Hirschfeld, L. A. (2001). On a folk theory of society: Children, evolution, and mental
12
13 representations of social groups. *Personality and Social Psychology Review*, *5*(2), 107-
14
15 117.

16
17
18
19 Hoppenstand, G. (1998). *Popular fiction: An anthology*. London, UK: Longman.

20
21 Humphrey, N. (1984). *Consciousness regained*. Oxford, UK: Oxford University Press.

22
23
24 Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*,
25
26 *342*(6156), 377-380.

27
28
29 Knox, J. (2003). *Archetype, attachment, analysis: Jungian psychology and the emergent mind*.
30
31 London, UK: Brunner-Routledge.

32
33 Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about
34
35 others. *Annual Review of Psychology*, *51*(1), 93-120.

36
37
38 Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworm versus nerds:
39
40 Exposure to fiction versus non-fiction, divergent associations with personality, ability,
41
42 and achievement. *Journal of Research in Personality*, *40*(5), 694-712.

43
44
45 Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of
46
47 social experience. *Perspectives on Psychological Science*, *3*(3), 173-192.

48
49
50 Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and
51
52 empathy: Ruling out individual differences and examining outcomes. *Communications*,
53
54 *34*(4), 407-428.

55
56
57
58 Mascaro, J. S., Rilling, J. K., Negi, L. T., & Raison, C. (2012). Compassion meditation enhances
59
60
61
62
63
64
65

- empathic accuracy and related neural activity. *Social Cognitive and Affective Neuroscience*, 8(1), 48-55.
- Maskit, B., & Murphy, S. (2014). The Discourse Attributes Analysis Program (DAAP). Retrieved September 01, 2014, from <http://www.thereferentialprocess.org/the-discourseattributes-analysis-program-daap>
- Mergenthaler, E., & Bucci, W. (1999). Linking verbal and non-verbal representations: Computer analysis of referential activity. *British Journal of Medical Psychology*, 72(3), 339-354.
- Miall, D. S., & Kuiken, D. (1999). What is literariness? Three components of literary reading. *Discourse Processes*, 28(2), 121-138.
- Mitchell, R., & Phillips, L. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, 70, 1-10.
- Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, 53(3), 431.
- Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).
- Schurz, M., & Perner, J. (2015). An evaluation of neurocognitive models of theory of mind. *Frontiers in Psychology*, 6.
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, 30(6), 855-863.
- Slade, A. (2005). Parental reflective functioning: An introduction. *Attachment & Human Development*, 7(3), 269-281.

- 1
2
3
4 Tamir, D., Bricker, A., Dodell-Feder, D., & Mitchell, J. (2016). Reading fiction and reading
5
6 minds: The role of simulation in the default network. *Social Cognitive and Affective*
7
8 *Neuroscience, 11(2)*, 215-224.
9
- 10
11 Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A.
12
13 (2013). The “Reading the Mind in the Eyes” test: Systematic review of psychometric
14
15 properties and a validation study in Italy. *Cognitive Neuropsychiatry, 18(4)*, 326-354.
16
17
- 18
19 Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice social-cognitive goals affect
20
21 amygdala and stereotype activation. *Psychological Science, 16(1)*, 56-63.
22
23
- 24
25 Wiessner, P. W. (2014). Embers of society: Firelight talk among the Ju/'hoansi Bushmen.
26
27 *Proceedings of the National Academy of Sciences, 111(39)*, 14027-14035.
28
- 29
30 Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation
31
32 in face recognition: A critical review of theories of the cross-race effect. *Personality and*
33
34 *Social Psychology Review, 16(2)*, 116-142.
35
- 36
37 Zunshine, L. (2015). From the social to the literary: Approaching Cao Xueqin's *The Story of the*
38
39 *Stone* from a cognitive perspective. In L. Zunshine (Ed.), *The Oxford handbook of*
40
41 *cognitive literary studies* (176-196). New York, NY: Oxford University Press.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Appendix A

Table of Authors, Texts, and RF Scores

Author	Title	Type	Delta RF
Berry, Wendell	Nothing Living Lives Alone (2011)*	Literary	.3100
Chekhov, Anton	Chameleon (1884)	Literary	.0751
Christie, Agatha	Tuesday Night Club (1932)*	Genre	.2004
Clancy, Tom	Command Authority (2013)*	Genre	.2227
Davis, Lydia	Blind Date (2009)	Literary	.1838
DeLillo, Don	The Runner (1988)	Literary	.2408
Drury, Tom	Pacific (2013)*	Literary	.1515
Gilb, Dagoberto	Uncle Rock (2010)*	Literary	.2539
Graver, Elizabeth	The End of the Point (2013)*	Literary	.2259
Hammett, Dashiell	Too Many Have Lived (1932)*	Genre	.1136
Heinlein, Robert	Space Jockey (1947)*	Genre	.1019
Lewis, R.J.	Ignite (2013)*	Genre	.1273
Lynn, J.	Be With Me (2014)*	Genre	.1632
Mattison, Alice	The Vandercook (2011)*	Literary	.1581
Maynard, Joyce	Labor Day (2009)*	Genre	.1193
McBride, James	The Good Lord Bird (2013)*	Literary	.1903
McDermott, Alice	Someone (2013)*	Literary	.2388
Munro, Alice	Corrie (2010)*	Literary	.2319
Pilcher, Rosamunde	Lalla (1990)*	Genre	.1700
Pynchon, Thomas	Bleeding Edge (2013)*	Literary	.1692
Riddle, A.G.	The Atlantis Gene (2013)*	Genre	.2536
Rinehart, Mary Roberts	Jane (1919)*	Genre	.2279
Ruddick, Sam	Leak (2010)*	Literary	.1351

Note: * Text used in present experiment

Appendix B

CRF Markers for High and Low Reflective Function

High RF Words	Low RF Words
and	I
was	to
that	you
the	my
of	t
a	s
it	me
she	so
know	they
he	just
in	would
but	when
think	don
her	no
with	m
had	not
very	then
sort	can
about	one
because	go

Note: Single-letter words (apart from “I”) stem from word contractions: for instance, “t” from “not” such as in “can’t.”