

Originally posted 3 October 2013; revised 4 October 2013



www.sciencemag.org/cgi/content/full/science.1239918/DC1

Supplementary Materials for  
**Reading Literary Fiction Improves Theory of Mind**

David Comer Kidd\* and Emanuele Castano\*

\*Corresponding author. E-mail: [kidd305@newschool.edu](mailto:kidd305@newschool.edu) (D.C.K.); [castano@newschool.edu](mailto:castano@newschool.edu) (E.C.)

Published 3 October 2013 on *Science Express*  
DOI: 10.1126/science.1239918

**This PDF file includes:**

Materials and Methods  
Supplementary Text  
Full References

**Correction:** Minor typographical errors.

## **Materials and Methods**

Because the experiments were conducted online, special attention was paid to indicators of attention and careless responding. In addition to participants with inadequately short reading times for the assigned texts, across all experiments, outliers (i.e.,  $> 3.5$  SD from the mean) on reading time, the Author Recognition Test guessing score, or on the dependent variables were excluded. Additionally, in Experiments 4 and 5, participants who reported having already participated in one of the experiments and participants for whom English was a second language were excluded (measures of these factors were not included in the first three experiments).

### Experiment 1

**Participants.** Participants ( $N = 90$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. One participant was excluded because of a high reading time ( $> 3.5$  SD), and two were excluded for high rates of guessing on the Author Recognition Test ( $> 3.5$  SD). There were no outliers on the dependent variable. An additional participant who did not read the assigned text (reading time of 0s) was excluded. The final sample included 86 participants (39 males), who ranged in age from 18 to 66 ( $M = 34.27$ ,  $SD = 11.94$ ).

**Materials and Procedure.** Six texts (3 fiction, 3 nonfiction) were selected by the authors. Critical to the selection were the criteria that the works of fiction depicted at least two characters and the nonfiction primarily focused on a nonhuman subject. These criteria were used to focus on the effects of reading about individuals presented in

literature compared to those of simply reading a well-written text. Two of the texts in the literary fiction condition, “The Runner” by Don DeLillo (38) and “Blind Date” by Lydia Davis (39), were written by contemporary award-winning authors. The third, “Chameleon”, was written by Anton Chekhov (40), an early master of the modern short story. The nonfiction texts were “How the Potato Changed the World” by Charles C. Mann (41), “Bamboo Steps Up” by Cathie Gandel (42), and “The Story of the Most Common Bird in the World” by Rob Dunn (43). Participants in each condition were randomly assigned to read one of the three appropriate texts.

Participants then completed a classic measure of cognitive ToM (23). They were presented with a scenario in which a character, Vicki, places a violin in a blue box and leaves the room. Then, another character enters the room and switches the location of the violin to a red box (false belief scenario) or switches the locations of the blue and red boxes (no false belief scenario). Participants were asked to indicate the probability (as a percentage) of Vicki looking in the red box. In the false belief scenario, this probability indicates failure to form a clear representation of Vicki’s false belief and so provides a measure of cognitive ToM. Although typically used in developmental research with young people, this paradigm has been used successfully to detect temporary variation in ToM performance among normal adults (23).

Participants also completed the Reading the Mind in the Eyes Test (RMET), an advanced affective ToM task assessing accuracy in emotion perception (6). For each of this task’s 36 trials, participants were asked to choose which of four emotions was being expressed in an image of only the eyes of an actor. This test is widely used as a measure of affective ToM, and scores have been shown to predict autistic traits in normal

populations (6), as well as nonclinical psychopathic characteristics, Machiavellianism, and empathy (44).

Next came the Author Recognition Test (24), which correlates positively with performance on the RMET (10-12). For this task, participants are given a list of 130 names, 65 of which are those of authors, and asked to identify any authors they recognize. Scores are calculated by subtracting the number of non-authors selected (the guessing score) from the number of authors identified and were square root transformed to improve the normality of the distribution.

Finally, participants were asked to complete the Positive and Negative Affect Schedule (PANAS; 25), a single item assessing current sadness, the Transportation scale (26), which assesses absorption in the text, and a series of demographic questions, including their age, gender, ethnicity and highest level of attained education. An additional measure of social perception (i.e., individuation) was collected for exploratory purposes and is not discussed here.

## Experiment 2

**Participants.** Participants ( $N = 117$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Applying the criteria for exclusion outlined above resulted in the deletion of one participant due to a high reading and another based on a high rate of guessing on the Author Recognition Test. There were no outliers on the dependent variable. In addition, a participant who did not read the text ( $M = 1.13\text{s}/\text{page}$ ) was excluded. The final sample included 114 participants (56 males,  $M_{\text{age}} = 33.30$ ,  $SD = 10.38$ ).

**Materials and Procedure.** Excerpts of the first several pages (8-11) of recently published novels were used as stimuli, with the stipulation that excerpts did not end in the middle of a scene or paragraph. In the literary fiction condition, participants read an excerpt from one of three recent finalists for the National Book Award for fiction [*The Round House* by Louise Erdrich (45), *The Tiger's Wife* by Téa Obreht (46), and *Salvage the Bones* by Jesmyn Ward (47)]. In the popular fiction condition, participants read an excerpt from one of three recent bestsellers on Amazon [*Gone Girl* by Gillian Flynn (48), *The Sins of the Mother* by Danielle Steel (49), and *Cross Roads* by W. Paul Young (50)]. Participants in the control condition read no text.

To assess the generality of the effects of reading literature on affective ToM, we used a different measure of affective ToM: the Diagnostic Analysis of Nonverbal Accuracy adult faces test (DANVA2-AF; 27). The DANVA2-AF has been extensively validated with normal, clinical, adult, and child populations, and predicts outcomes related to competence in social interactions, ranging from psychopathological symptoms to occupational success (27). The DANVA2-AF includes images of actors' entire faces as they express anger, fear, sadness, or happiness at high or low levels of intensity. Faces were shown for 2s, followed by a mask, and participants were asked to choose the emotion term corresponding to the expressed emotion. As in previous research (51), the happiness subscale of the DANVA2-AF correlated negatively with the other subscales (sadness, anger and fear), and removing it substantially increased internal consistency (Cronbach's  $\alpha$  increased from  $\alpha = .53$  to  $\alpha = .65$ ).

The false-belief ToM task and the Author Recognition Test were used as in Experiment 1. Participants who were assigned texts also filled out the Transportation

scale. All participants reported their age, gender, and highest level of attained education. The same measure of individuation used in Experiment 1 was used in Experiment 2 for exploratory purposes, but it is not discussed further.

### Experiment 3

**Participants.** Participants ( $N = 78$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. While in Experiments 1 and 2 texts were presented in their original formats, in Experiment 3 the texts were retyped to standardize page length. This allowed for the adoption of a single criterion for excluding participants based on low reading times (i.e.,  $< 30$ s/page) that might not emerge as statistical outliers. This strategy was adopted in all subsequent experiments as well. Based on this criterion, 9 participants were excluded. Applying the exact same criteria used to identify outliers in the previous experiments resulted in no further exclusions. The final sample included 69 participants (40 males), who ranged in age from 18 to 62 ( $M = 33.88$ ,  $SD = 11.61$ ).

**Materials and Procedures.** Six new texts, 3 in each condition, were used. The stories in the popular fiction condition were selected from an edited anthology of popular fiction (29). They were also chosen to represent a range of genres, including science fiction [*Space Jockey* by Robert Heinlein], mystery [*Too Many Have Lived* by Dashiell Hammett] and romance [*Lalla* by Rosamunde Pilcher]. Stories in the literary fiction condition were selected from a collection of the 2012 winners of the PEN/O. Henry Award for short literary fiction (30). They included *Corrie* by Alice Munro, *Leak* by Sam Ruddick, and *Nothing Living Lives Alone* by Wendell Berry.

After reading their assigned story, participants completed the RMET and the Author Recognition Test. All participants also completed the PANAS, answered a single item about current sadness, and reported their age, gender, ethnicity, and highest attained education.

#### Experiment 4

**Participants.** Participants ( $N = 106$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Because participants were recruited from the same pool as in the previous experiments, participants in this experiment (and Experiment 5) were asked whether they had participated in one of the prior experiments. Those who reported having done so were excluded ( $n = 13$ ). In this experiment (and Experiment 5), for the first time, participants also indicated their native language. One participant, who was not a native English speaker, was excluded. As in Experiment 3, texts were retyped, and applying the same criterion for excluding participants with low reading times (i.e.,  $< 30$ s) resulted in the exclusion of 14 participants. The exact same criteria as in the prior three experiments were used to delete outliers: one on the Author Recognition Test; five on error rates on the control trials of the Yoni task (described below); no participants on the RMET or on the critical Yoni trials. The final sample included 72 participants (20 males), who ranged in age from 18 to 59 ( $M = 34.26$ ,  $SD = 10.90$ ).

**Materials and Procedures.** Four of the texts used in Experiment 4 were the same as those used in Experiment 3. Two new texts, *Jane* by Mary Roberts Rinehart (29, popular fiction) and *Uncle Rock* by Dagoberto Gilb (30, literary fiction), replaced *Lalla* (29) and *Leak* (30) from Experiment 3. After reading their assigned text, participants

completed the RMET followed by the Yoni task, which assesses both cognitive and affective ToM (7). Finally, participants completed the Author Recognition Test and answered demographic questions including age, gender, and level of education.

The Yoni task includes first order (tracking one target's mental state) and second order (tracking two targets' mental states) cognitive (24 trials) and affective (24 trials) sections, as well as 14 control trials in which participants must only identify physical characteristics of images. For each trial, participants are shown four images surrounding a central character (usually named Yoni, but called John in this study). For the cognitive ToM trials, participants must identify, based on visual and linguistic cues, which of the four surrounding images the central character is thinking about or wants. For affective ToM trials, participants must make similar decisions about which image the central character likes, loves, dislikes, or does not love. Scores reflect proportion of correct responses. This task has been used in a wide range of populations to assess both components of ToM (7, 8, 31).

The RMET was administered in the same way as in Experiments 1 and 3.

**Yoni Task Analysis.** Performance on the Yoni task was analyzed using a mixed analysis of covariance (ANCOVA) with type of trial (cognitive, affective) and trial difficulty (first order, second order) as within-participants factors. Experimental condition (literary fiction, popular fiction) and Author Recognition Test were between-participants factors, and performance on control trials was a covariate. Aside from the effects discussed in the main text, there was a main effect of type of trial [ $F(1,67) = 6.08$ ,  $P = .01$ ], trial difficulty [ $F(1,67) = 6.43$ ,  $P = .01$ ], and performance on control trials [ $\beta = 0.49$ ,  $P < .001$ ]. There were significant interactions of trial difficulty and trial type

[ $F(1,67) = 9.98, P = .002$ ] and of trial type and performance on control trials [ $F(1,67) = 5.35, P = .02$ ]. In addition, there was a three-way interaction of type of trial, trial difficulty, and performance on control trials [ $F(1,67) = 8.15, P = .005$ ].

### Experiment 5

**Participants.** Participants ( $N = 456$ ) were recruited and compensated (\$3.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Participants ( $n = 22$ ) who had completed one of the previous experiments were excluded. Participants who reported not being native English speakers were excluded ( $n = 9$ ). As in the previous two experiments page length was standardized and participants who spent less than 30 seconds per page of text were also excluded ( $n = 29$ ). The exact same criteria as in the prior four experiments were used to delete outliers: five on reading time; 24 on the Author Recognition Test guessing score; 9 on error rates on the control trials of the Yoni task; one on the RMET; 1 on the critical trials of the Yoni task. The final sample included 356 participants (182 males), who ranged in age from 18 to 75 ( $M = 34.42, SD = 11.59$ ).

**Materials and Procedures.** Five of the texts used in Experiment 5 were the same as those used in Experiment 4, but *The Vandercook* by Alice Mattinson (30) replaced *Nothing Alive Lives Alone* by Wendell Berry (30) in the literary fiction condition because it was shorter and so closer in length to the other texts. Participants completed the RMET followed by the Yoni task. Then, participants completed the PANAS and two items assessing current feelings of sadness and happiness on 7-point scales before completing the Author Recognition Test. Participants in the reading conditions then completed the Transportation scale and responded to two items assessing how much they enjoyed the

text and how much they thought it was “an example of excellent literature” on 7-point scales. Finally, all participants answered demographic questions assessing their age, gender, ethnicity, and highest level of attained education.

**Yoni Task Analysis.** As in Experiment 4, performance on the Yoni task was analyzed using a mixed ANCOVA with type of trial (cognitive, affective) and trial difficulty (first order, second order) as within-participants factors. Experimental condition (literary fiction, popular fiction, no-reading) and Author Recognition Test scores were entered as between-participants factors, and performance on control trials as covariate. Aside from the effects discussed in the main text, there was a main effect of difficulty of trial [ $F(1,351) = 28.99, P < .001$ ] and a marginal effect of trial type [ $F(1,351) = 3.15, P = .07$ ]. The effect of performance on control trials was significant [ $F(1,351) = 25.40, P < .001$ ]. The interaction of trial difficulty and type of trial was significant [ $F(1,351) = 7.95, P = .005$ ], as was the interaction of trial difficulty and performance on control trials [ $F(1,351) = 8.59, P = .003$ ].

## References

1. C. Batson, in *The Handbook of Social Psychology*, D.T. Gilbert, S.T. Fiske, G. Lindzey, Eds. (McGraw-Hill, Boston, 1998), pp. 282–316.
2. F. B. de Waal, The antiquity of empathy. *Science* **336**, 874–876 (2012). [Medline doi:10.1126/science.1220999](https://doi.org/10.1126/science.1220999)
3. R. Saxe, S. Carey, N. Kanwisher, Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.* **55**, 87–124 (2004). [Medline doi:10.1146/annurev.psych.55.090902.142044](https://doi.org/10.1146/annurev.psych.55.090902.142044)
4. C. L. Sebastian, N. M. G. Fontaine, G. Bird, S.-J. Blakemore, S. A. De Brito, E. J. P. McCrory, E. Viding, Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Soc. Cogn. Affect. Neurosci.* **7**, 53–63 (2012). [Medline doi:10.1093/scan/nsr023](https://doi.org/10.1093/scan/nsr023)

5. S. G. Shamay-Tsoory, The neural bases for empathy. *Neuroscientist* **17**, 18–24 (2011). [Medline doi:10.1177/1073858410379268](#)
6. S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, The “Reading the Mind in the Eyes” Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* **42**, 241–251 (2001). [Medline doi:10.1111/1469-7610.00715](#)
7. S. G. Shamay-Tsoory, J. Aharon-Peretz, Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia* **45**, 3054–3067 (2007). [Medline doi:10.1016/j.neuropsychologia.2007.05.021](#)
8. S. G. Shamay-Tsoory, H. Harari, J. Aharon-Peretz, Y. Levkovitz, The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex* **46**, 668–677 (2010). [Medline doi:10.1016/j.cortex.2009.04.008](#)
9. E. Castano, in *The Oxford Handbook of Personality and Social Psychology*, K. Deaux and M. Snyder, Eds. (Oxford Univ. Press, New York, 2012), pp. 419–445.
10. R. Mar, K. Oatley, J. Hirsh, J. dela Paz, J. Peterson, Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *J. Res. Pers.* **40**, 694–712 (2006). [doi:10.1016/j.jrp.2005.08.002](#)
11. R. Mar, K. Oatley, J. Peterson, Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications* **34**, 407–428 (2009). [doi:10.1515/COMM.2009.025](#)
12. M. Djikic, K. Oatley, M. Moldoveanu, Reading other minds: Effects of literature on empathy. *Scientific Study of Literature* **3**, 28–47 (2013). [doi:10.1075/ssol.3.1.06dji](#)
13. P. M. Bal, M. Veltkamp, How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS ONE* **8**, e55341 (2013). [Medline doi:10.1371/journal.pone.0055341](#)
14. E. Schiappa, P. Gregg, D. Hewes, The parasocial contact hypothesis. *Commun. Monogr.* **72**, 92–115 (2005). [doi:10.1080/0363775052000342544](#)
15. D. Miall, D. Kuiken, Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics* **22**, 389–407 (1994). [doi:10.1016/0304-422X\(94\)00011-5](#)
16. D. Miall, D. Kuiken, Beyond text theory: Understanding literary response. *Discourse Process.* **17**, 337–352 (1994). [doi:10.1080/01638539409544873](#)
17. D. Miall, D. Kuiken, What is literariness? Three components of literary reading. *Discourse Process.* **28**, 121–138 (1999). [doi:10.1080/01638539909545076](#)
18. R. Barthes, *S/Z: An Essay* (Hill and Wang, New York, 1974).
19. M. Bakhtin, *Problems of Dostoevsky's Poetics* (Univ. Minnesota, Minneapolis, 1984).
20. J. Bruner, *Actual Minds, Possible Worlds* (Harvard Univ., Cambridge, MA, 1986).

21. R. Gerrig, D. Rapp, Psychological processes underlying literary impact. *Poetics Today* **25**, 265–281 (2004). [doi:10.1215/03335372-25-2-265](https://doi.org/10.1215/03335372-25-2-265)
22. Materials and methods are available as supplementary materials on *Science Online*.
23. B. A. Converse, S. Lin, B. Keysar, N. Epley, In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion* **8**, 725–730 (2008). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.1037/a0013283](https://doi.org/10.1037/a0013283)
24. D. J. Acheson, J. B. Wells, M. C. MacDonald, New and updated tests of print exposure and reading abilities in college students. *Behav. Res. Methods* **40**, 278–289 (2008). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.3758/BRM.40.1.278](https://doi.org/10.3758/BRM.40.1.278)
25. D. Watson, L. A. Clark, A. Tellegen, Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.1037/0022-3514.54.6.1063](https://doi.org/10.1037/0022-3514.54.6.1063)
26. M. C. Green, T. C. Brock, The role of transportation in the persuasiveness of public narratives. *J. Pers. Soc. Psychol.* **79**, 701–721 (2000). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.1037/0022-3514.79.5.701](https://doi.org/10.1037/0022-3514.79.5.701)
27. S. Nowicki, *Manual for the Receptive Tests of the Diagnostic Analysis of Nonverbal Accuracy 2* (Department of Psychology, Emory Univ., Atlanta, GA, 2010).
28. R. Mar, K. Oatley, The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* **3**, 173–192 (2008). [doi:10.1111/j.1745-6924.2008.00073.x](https://doi.org/10.1111/j.1745-6924.2008.00073.x)
29. G. Hoppenstand, Ed., *Popular Fiction: An Anthology* (Longman, New York, 1998).
30. L. Furman, Ed., *The PEN/O. Henry Prize Stories 2012: The Best Stories of the Year* (Anchor, New York, 2012).
31. S. G. Shamay-Tsoory, Recognition of ‘fortune of others’ emotions in Asperger syndrome and high functioning autism. *J. Autism Dev. Disord.* **38**, 1451–1461 (2008). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.1007/s10803-007-0515-9](https://doi.org/10.1007/s10803-007-0515-9)
32. B. J. Bushman, C. A. Anderson, Comfortably numb: Desensitizing effects of violent media on helping others. *Psychol. Sci.* **20**, 273–277 (2009). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.1111/j.1467-9280.2009.02287.x](https://doi.org/10.1111/j.1467-9280.2009.02287.x)
33. J. Pennebaker, R. Booth, M. Francis, *Linguistic Inquiry and Word Count (LIWC): LIWC2007* (LIWC.net, Austin, TX, 2007).
34. N. Humphrey, *Consciousness Regained* (Oxford Univ., Oxford, 1983).
35. M. F. McLellan, A. H. Jones, Why literature and medicine? *Lancet* **348**, 109–111 (1996). [Medline](https://pubmed.ncbi.nlm.nih.gov/18111111/) [doi:10.1016/S0140-6736\(96\)03521-0](https://doi.org/10.1016/S0140-6736(96)03521-0)
36. J. Billington, ‘Reading for Life’: Prison reading groups in practice and theory. *Critical Survey* **23**, 67–85 (2011). [doi:10.3167/cs.2011.230306](https://doi.org/10.3167/cs.2011.230306)
37. S. Mosle, What should children read? *New York Times*, 22 November 2012; <http://opinionator.blogs.nytimes.com/2012/11/22/what-should-children-read/?r=1>.
38. D. DeLillo, *The Angel Esmeralda* (Scribner, New York, 2011), pp. 47–54.

39. L. Davis, *The Collected Stories of Lydia Davis* (Farrar, Straus, & Giroux, New York, 2009), pp. 333–337.
40. A. Chekhov, in *Anton Chekhov's Short Stories*, R. E. Matlow, Ed. (Norton, New York, 1979; original work published 1884), pp. 1–3.
41. C. Mann, How the potato changed the world. *Smithsonian Magazine*, November 2011; [www.smithsonianmag.com/history-archaeology/How-the-Potato-Changed-the-World.html](http://www.smithsonianmag.com/history-archaeology/How-the-Potato-Changed-the-World.html).
42. C. Gandel, Bamboo steps up. *Smithsonian Magazine*, 21 March 2008; [www.smithsonianmag.com/specialsections/ecocenter/greener-living/bamboo.html](http://www.smithsonianmag.com/specialsections/ecocenter/greener-living/bamboo.html).
43. R. Dunn, The story of the most common bird in the world. *Smithsonian Magazine*, 2 March 2012; [www.smithsonianmag.com/science-nature/The-Story-of-the-Most-Common-Bird-in-the-World.html](http://www.smithsonianmag.com/science-nature/The-Story-of-the-Most-Common-Bird-in-the-World.html).
44. F. Ali, T. Chamorro-Premuzic, Investigating theory of mind deficits in nonclinical psychopathy and Machiavellianism. *Pers. Individ. Dif.* **49**, 169–174 (2010). [doi:10.1016/j.paid.2010.03.027](https://doi.org/10.1016/j.paid.2010.03.027)
45. L. Erdrich, *The Round House* (Harper Collins, New York, 2012).
46. T. Obrecht, *The Tiger's Wife* (Random House, New York, 2011).
47. J. Ward, *Salvage the Bones* (Bloomsbury, New York, 2011).
48. G. Flynn, *Gone Girl* (Crown Publishers, New York, 2012).
49. D. Steele, *The Sins of the Mother* (Random House, New York, 2012).
50. W. Young, *Cross Roads* (FaithWords, Nashville, TN, 2012).
51. H. A. Elfenbein, N. Ambady, Predicting workplace outcomes from the ability to eavesdrop on feelings. *J. Appl. Psychol.* **87**, 963–971 (2002). [Medline](https://pubmed.ncbi.nlm.nih.gov/12111111/) [doi:10.1037/0021-9010.87.5.963](https://doi.org/10.1037/0021-9010.87.5.963)

Originally posted 3 October 2013; revised 4 October 2013



# Supplementary Materials for Reading Literary Fiction Improves Theory of Mind

David Comer Kidd and Emanuele Castano

correspondence to:

**This PDF file includes:**

Materials and Methods  
Supplementary Text

Originally posted 3 October 2013; revised 4 October 2013

## **Materials and Methods:**

Since the experiments were conducted online, special attention was paid to indicators of attention and careless responding. In addition to participants with inadequately short reading times for the assigned texts, across all experiments, outliers (i.e.,  $>3.5 SD$  from the mean) on reading time, the Author Recognition Test guessing score, or on the dependent variables were excluded. Additionally, in Experiments 4 and 5, participants who reported having already participated in one of the experiments and participants for whom English was a second language were excluded (measures of these factors were not included in the first three experiments).

### Experiment 1

**Participants.** Participants ( $N = 90$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. One participant was excluded because of a high reading time ( $>3.5 SD$ ), and two were excluded for high rates of guessing on the Author Recognition Test ( $>3.5 SD$ ). There were no outliers on the dependent variable. An additional participant who did not read the assigned text (reading time of 0s) was excluded. The final sample included 86 participants (39 males), who ranged in age from 18 to 66 ( $M = 34.27$ ,  $SD = 11.94$ ).

**Materials and Procedure.** Six texts (3 fiction, 3 nonfiction) were selected by the authors. Critical to the selection were the criteria that the works of fiction depicted at least two characters and the nonfiction primarily focused on a nonhuman subject. These criteria were used to focus on the effects of reading about individuals presented in

Originally posted 3 October 2013; revised 4 October 2013

literature compared to those of simply reading a well-written text. Two of the texts in the literary fiction condition, “The Runner” by Don DeLillo (38) and “Blind Date” by Lydia Davis (39), were written by contemporary award-winning authors. The third, “Chameleon”, was written by Anton Chekhov (40), an early master of the modern short story. The nonfiction texts were “How the Potato Changed the World” by Charles C. Mann (41), “Bamboo Steps Up” by Cathie Gandel (42), and “The Story of the Most Common Bird in the World” by Rob Dunn (43). Participants in each condition were randomly assigned to read one of the three appropriate texts.

Participants then completed a classic measure of cognitive ToM (23). They were presented with a scenario in which a character, Vicki, places a violin in a blue box and leaves the room. Then, another character enters the room and switches the location of the violin to a red box (false belief scenario) or switches the locations of the blue and red boxes (no false belief scenario). Participants were asked to indicate the probability (as a percentage) of Vicki looking in the red box. In the false belief scenario, this probability indicates failure to form a clear representation of Vicki’s false belief and so provides a measure of cognitive ToM. Although typically used in developmental research with young people, this paradigm has been used successfully to detect temporary variation in ToM performance among normal adults (23).

Participants also completed the Reading the Mind in the Eyes Test (RMET), an advanced affective ToM task assessing accuracy in emotion perception (6). For each of this task’s 36 trials, participants were asked to choose which of four emotions was being expressed in an image of only the eyes of an actor. This test is widely used as a measure of affective ToM, and scores have been shown to predict autistic traits in normal

Originally posted 3 October 2013; revised 4 October 2013

populations (6), as well as nonclinical psychopathic characteristics, Machiavellianism, and empathy (44).

Next came the Author Recognition Test (24), which correlates positively with performance on the RMET (10-12). For this task, participants are given a list of 130 names, 65 of which are those of authors, and asked to identify any authors they recognize. Scores are calculated by subtracting the number of non-authors selected (the guessing score) from the number of authors identified and were square root transformed to improve the normality of the distribution.

Finally, participants were asked to complete the Positive and Negative Affect Schedule (PANAS; 25), a single item assessing current sadness, the Transportation scale (26), which assesses absorption in the text, and a series of demographic questions, including their age, gender, ethnicity and highest level of attained education. An additional measure of social perception (i.e., individuation) was collected for exploratory purposes and is not discussed here.

## Experiment 2

**Participants.** Participants ( $N = 117$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Applying the criteria for exclusion outlined above resulted in the deletion of one participant due to a high reading and another based on a high rate of guessing on the Author Recognition Test. There were no outliers on the dependent variable. In addition, a participant who did not read the text ( $M = 1.13\text{s}/\text{page}$ ) was excluded. The final sample included 114 participants (56 males,  $M_{\text{age}} = 33.30$ ,  $SD = 10.38$ ).

Originally posted 3 October 2013; revised 4 October 2013

**Materials and Procedure.** Excerpts of the first several pages (8-11) of recently published novels were used as stimuli, with the stipulation that excerpts did not end in the middle of a scene or paragraph. In the literary fiction condition, participants read an excerpt from one of three recent finalists for the National Book Award for fiction [*The Round House* by Louise Erdrich (45), *The Tiger's Wife* by Téa Obreht (46), and *Salvage the Bones* by Jesmyn Ward (47)]. In the popular fiction condition, participants read an excerpt from one of three recent bestsellers on Amazon [*Gone Girl* by Gillian Flynn (48), *The Sins of the Mother* by Danielle Steel (49), and *Cross Roads* by W. Paul Young (50)]. Participants in the control condition read no text.

To assess the generality of the effects of reading literature on affective ToM, we used a different measure of affective ToM: the Diagnostic Analysis of Nonverbal Accuracy adult faces test (DANVA2-AF; 27). The DANVA2-AF has been extensively validated with normal, clinical, adult, and child populations, and predicts outcomes related to competence in social interactions, ranging from psychopathological symptoms to occupational success (27). The DANVA2-AF includes images of actors' entire faces as they express anger, fear, sadness, or happiness at high or low levels of intensity. Faces were shown for 2s, followed by a mask, and participants were asked to choose the emotion term corresponding to the expressed emotion. As in previous research (51), the happiness subscale of the DANVA2-AF correlated negatively with the other subscales (sadness, anger and fear), and removing it substantially increased internal consistency (Cronbach's  $\alpha$  increased from  $\alpha = .53$  to  $\alpha = .65$ ).

The false-belief ToM task and the Author Recognition Test were used as in Experiment 1. Participants who were assigned texts also filled out the Transportation

Originally posted 3 October 2013; revised 4 October 2013

scale. All participants reported their age, gender, and highest level of attained education. The same measure of individuation used in Experiment 1 was used in Experiment 2 for exploratory purposes, but it is not discussed further.

### Experiment 3

**Participants.** Participants ( $N = 78$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. While in Experiments 1 and 2 texts were presented in their original formats, in Experiment 3 the texts were retyped to standardize page length. This allowed for the adoption of a single criterion for excluding participants based on low reading times (i.e.,  $< 30$ s/page) that might not emerge as statistical outliers. This strategy was adopted in all subsequent experiments as well. Based on this criterion, 9 participants were excluded. Applying the exact same criteria used to identify outliers in the previous experiments resulted in no further exclusions. The final sample included 69 participants (40 males), who ranged in age from 18 to 62 ( $M = 33.88$ ,  $SD = 11.61$ ).

**Materials and Procedures.** Six new texts, 3 in each condition, were used. The stories in the popular fiction condition were selected from an edited anthology of popular fiction (29). They were also chosen to represent a range of genres, including science fiction [*Space Jockey* by Robert Heinlein], mystery [*Too Many Have Lived* by Dashiell Hammett] and romance [*Lalla* by Rosamunde Pilcher]. Stories in the literary fiction condition were selected from a collection of the 2012 winners of the PEN/O. Henry Award for short literary fiction (30). They included *Corrie* by Alice Munroe, *Leak* by Sam Ruddick, and *Nothing Living Lives Alone* by Wendell Berry.

Originally posted 3 October 2013; revised 4 October 2013

After reading their assigned story, participants completed the RMET and the Author Recognition Test. All participants also completed the PANAS, answered a single item about current sadness, and reported their age, gender, ethnicity, and highest attained education.

#### Experiment 4

**Participants.** Participants ( $N = 106$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Since participants were recruited from the same pool as in the previous experiments, participants in this experiment (and Experiment 5) were asked whether they had participated in one of the prior experiments. Those who reported having done so were excluded ( $n = 13$ ). In this experiment (and Experiment 5), for the first time, participants also indicated their native language. One participant, who was not a native English speaker, was excluded. As in Experiment 3, texts were retyped, and applying the same criterion for excluding participants with low reading times (i.e.,  $< 30$ s) resulted in the exclusion of 14 participants. The exact same criteria as in the prior three experiments were used to delete outliers: one on the Author Recognition Test; five on error rates on the control trials of the Yoni task (described below); no participants on the RMET or on the critical Yoni trials. The final sample included 72 participants (20 males), who ranged in age from 18 to 59 ( $M = 34.26$ ,  $SD = 10.90$ ).

**Materials and Procedures.** Four of the texts used in Experiment 4 were the same as those used in Experiment 3. Two new texts, *Jane* by Mary Jane Rinehart (29, popular fiction) and *Uncle Rock* by Dagoberto Gilb (30, literary fiction), replaced *Lalla* (29) and *Leak* (30) from Experiment 3. After reading their assigned text, participants completed

Originally posted 3 October 2013; revised 4 October 2013

the RMET followed by the Yoni task, which assesses both cognitive and affective ToM (7). Finally, participants completed the Author Recognition Test and answered demographic questions including age, gender, and level of education.

The Yoni task includes first order (tracking one target's mental state) and second order (tracking two targets' mental states) cognitive (24 trials) and affective (24 trials) sections, as well as 14 control trials in which participants must only identify physical characteristics of images. For each trial, participants are shown four images surrounding a central character (usually named Yoni, but called John in this study). For the cognitive ToM trials, participants must identify, based on visual and linguistic cues, which of the four surrounding images the central character is thinking about or wants. For affective ToM trials, participants must make similar decisions about which image the central character likes, loves, dislikes, or does not love. Scores reflect proportion of correct responses. This task has been used in a wide range of populations to assess both components of ToM (7, 8, 31).

The RMET was administered in the same way as in Experiments 1 and 3.

### **Yoni Task Analysis**

Performance on the Yoni task was analyzed using a mixed analysis of covariance (ANCOVA) with type of trial (cognitive, affective) and trial difficulty (first order, second order) as within-participants factors. Experimental condition (literary fiction, popular fiction) and Author Recognition Test were between-participants factors, and performance on control trials was a covariate. Aside from the effects discussed in the main text, there was a main effect of type of trial [ $F(1,67) = 6.08, P = .01$ ], trial difficulty [ $F(1,67) = 6.43, P = .01$ ], and performance on control trials [ $\beta = 0.49, P < .001$ ]. There were significant

Originally posted 3 October 2013; revised 4 October 2013

interactions of trial difficulty and trial type [ $F(1,67) = 9.98, P = .002$ ] and of trial type and performance on control trials [ $F(1,67) = 5.35, P = .02$ ]. In addition, there was a three-way interaction of type of trial, trial difficulty, and performance on control trials [ $F(1,67) = 8.15, P = .005$ ].

### Experiment 5

**Participants.** Participants ( $N = 456$ ) were recruited and compensated (\$3.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Participants ( $n = 22$ ) who had completed one of the previous experiments were excluded. Participants who reported not being native English speakers were excluded ( $n = 9$ ). As in the previous two experiments page length was standardized and participants who spent less than 30 seconds per page of text were also excluded ( $n = 29$ ). The exact same criteria as in the prior four experiments were used to delete outliers: five on reading time; 24 on the Author Recognition Test guessing score; 9 on error rates on the control trials of the Yoni task; one on the RMET; 1 on the critical trials of the Yoni task. The final sample included 356 participants (182 males), who ranged in age from 18 to 75 ( $M = 34.42, SD = 11.59$ ).

**Materials and Procedures.** Five of the texts used in Experiment 5 were the same as those used in Experiment 4, but *The Vandercook* by Alice Mattinson (30) replaced *Nothing Alive Lives Alone* by Wendell Berry (30) in the literary fiction condition because it was shorter and so closer in length to the other texts. Participants completed the RMET followed by the Yoni task. Then, participants completed the PANAS and two items assessing current feelings of sadness and happiness on 7-point scales before completing the Author Recognition Test. Participants in the reading conditions then completed the

Originally posted 3 October 2013; revised 4 October 2013

Transportation scale and responded to two items assessing how much they enjoyed the text and how much they thought it was “an example of excellent literature” on 7-point scales. Finally, all participants answered demographic questions assessing their age, gender, ethnicity, and highest level of attained education.

### **Yoni Task Analysis**

As in Experiment 4, performance on the Yoni task was analyzed using a mixed ANCOVA with type of trial (cognitive, affective) and trial difficulty (first order, second order) as within-participants factors. Experimental condition (literary fiction, popular fiction, no-reading) and Author Recognition Test scores were entered as between-participants factors, and performance on control trials as covariate. Aside from the effects discussed in the main text, there was a main effect of difficulty of trial [ $F(1,351) = 28.99$ ,  $P < .001$ ] and a marginal effect of trial type [ $F(1,351) = 3.15$ ,  $P = .07$ ]. The effect of performance on control trials was significant [ $F(1,351) = 25.40$ ,  $P < .001$ ]. The interaction of trial difficulty and type of trial was significant [ $F(1,351) = 7.95$ ,  $P = .005$ ], as was the interaction of trial difficulty and performance on control trials [ $F(1,351) = 8.59$ ,  $P = .003$ ].

Originally posted 3 October 2013; revised 4 October 2013



# Supplementary Materials for Reading Literary Fiction Improves Theory of Mind

David Comer Kidd and Emanuele Castano

correspondence to:

**This PDF file includes:**

Materials and Methods  
Supplementary Text

Originally posted 3 October 2013; revised 4 October 2013

## **Materials and Methods:**

Since the experiments were conducted online, special attention was paid to indicators of attention and careless responding. In addition to participants with inadequately short reading times for the assigned texts, across all experiments, outliers (i.e.,  $>3.5 SD$  from the mean) on reading time, the Author Recognition Test guessing score, or on the dependent variables were excluded. Additionally, in Experiments 4 and 5, participants who reported having already participated in one of the experiments and participants for whom English was a second language were excluded (measures of these factors were not included in the first three experiments).

### Experiment 1

**Participants.** Participants ( $N = 90$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. One participant was excluded because of a high reading time ( $>3.5 SD$ ), and two were excluded for high rates of guessing on the Author Recognition Test ( $>3.5 SD$ ). There were no outliers on the dependent variable. An additional participant who did not read the assigned text (reading time of 0s) was excluded. The final sample included 86 participants (39 males), who ranged in age from 18 to 66 ( $M = 34.27$ ,  $SD = 11.94$ ).

**Materials and Procedure.** Six texts (3 fiction, 3 nonfiction) were selected by the authors. Critical to the selection were the criteria that the works of fiction depicted at least two characters and the nonfiction primarily focused on a nonhuman subject. These criteria were used to focus on the effects of reading about individuals presented in

Originally posted 3 October 2013; revised 4 October 2013

literature compared to those of simply reading a well-written text. Two of the texts in the literary fiction condition, “The Runner” by Don DeLillo (38) and “Blind Date” by Lydia Davis (39), were written by contemporary award-winning authors. The third, “Chameleon”, was written by Anton Chekhov (40), an early master of the modern short story. The nonfiction texts were “How the Potato Changed the World” by Charles C. Mann (41), “Bamboo Steps Up” by Cathie Gandel (42), and “The Story of the Most Common Bird in the World” by Rob Dunn (43). Participants in each condition were randomly assigned to read one of the three appropriate texts.

Participants then completed a classic measure of cognitive ToM (23). They were presented with a scenario in which a character, Vicki, places a violin in a blue box and leaves the room. Then, another character enters the room and switches the location of the violin to a red box (false belief scenario) or switches the locations of the blue and red boxes (no false belief scenario). Participants were asked to indicate the probability (as a percentage) of Vicki looking in the red box. In the false belief scenario, this probability indicates failure to form a clear representation of Vicki’s false belief and so provides a measure of cognitive ToM. Although typically used in developmental research with young people, this paradigm has been used successfully to detect temporary variation in ToM performance among normal adults (23).

Participants also completed the Reading the Mind in the Eyes Test (RMET), an advanced affective ToM task assessing accuracy in emotion perception (6). For each of this task’s 36 trials, participants were asked to choose which of four emotions was being expressed in an image of only the eyes of an actor. This test is widely used as a measure of affective ToM, and scores have been shown to predict autistic traits in normal

Originally posted 3 October 2013; revised 4 October 2013

populations (6), as well as nonclinical psychopathic characteristics, Machiavellianism, and empathy (44).

Next came the Author Recognition Test (24), which correlates positively with performance on the RMET (10-12). For this task, participants are given a list of 130 names, 65 of which are those of authors, and asked to identify any authors they recognize. Scores are calculated by subtracting the number of non-authors selected (the guessing score) from the number of authors identified and were square root transformed to improve the normality of the distribution.

Finally, participants were asked to complete the Positive and Negative Affect Schedule (PANAS; 25), a single item assessing current sadness, the Transportation scale (26), which assesses absorption in the text, and a series of demographic questions, including their age, gender, ethnicity and highest level of attained education. An additional measure of social perception (i.e., individuation) was collected for exploratory purposes and is not discussed here.

## Experiment 2

**Participants.** Participants ( $N = 117$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Applying the criteria for exclusion outlined above resulted in the deletion of one participant due to a high reading and another based on a high rate of guessing on the Author Recognition Test. There were no outliers on the dependent variable. In addition, a participant who did not read the text ( $M = 1.13\text{s}/\text{page}$ ) was excluded. The final sample included 114 participants (56 males,  $M_{\text{age}} = 33.30$ ,  $SD = 10.38$ ).

Originally posted 3 October 2013; revised 4 October 2013

**Materials and Procedure.** Excerpts of the first several pages (8-11) of recently published novels were used as stimuli, with the stipulation that excerpts did not end in the middle of a scene or paragraph. In the literary fiction condition, participants read an excerpt from one of three recent finalists for the National Book Award for fiction [*The Round House* by Louise Erdrich (45), *The Tiger's Wife* by Téa Obreht (46), and *Salvage the Bones* by Jesmyn Ward (47)]. In the popular fiction condition, participants read an excerpt from one of three recent bestsellers on Amazon [*Gone Girl* by Gillian Flynn (48), *The Sins of the Mother* by Danielle Steel (49), and *Cross Roads* by W. Paul Young (50)]. Participants in the control condition read no text.

To assess the generality of the effects of reading literature on affective ToM, we used a different measure of affective ToM: the Diagnostic Analysis of Nonverbal Accuracy adult faces test (DANVA2-AF; 27). The DANVA2-AF has been extensively validated with normal, clinical, adult, and child populations, and predicts outcomes related to competence in social interactions, ranging from psychopathological symptoms to occupational success (27). The DANVA2-AF includes images of actors' entire faces as they express anger, fear, sadness, or happiness at high or low levels of intensity. Faces were shown for 2s, followed by a mask, and participants were asked to choose the emotion term corresponding to the expressed emotion. As in previous research (51), the happiness subscale of the DANVA2-AF correlated negatively with the other subscales (sadness, anger and fear), and removing it substantially increased internal consistency (Cronbach's  $\alpha$  increased from  $\alpha = .53$  to  $\alpha = .65$ ).

The false-belief ToM task and the Author Recognition Test were used as in Experiment 1. Participants who were assigned texts also filled out the Transportation

Originally posted 3 October 2013; revised 4 October 2013

scale. All participants reported their age, gender, and highest level of attained education. The same measure of individuation used in Experiment 1 was used in Experiment 2 for exploratory purposes, but it is not discussed further.

### Experiment 3

**Participants.** Participants ( $N = 78$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. While in Experiments 1 and 2 texts were presented in their original formats, in Experiment 3 the texts were retyped to standardize page length. This allowed for the adoption of a single criterion for excluding participants based on low reading times (i.e.,  $< 30$ s/page) that might not emerge as statistical outliers. This strategy was adopted in all subsequent experiments as well. Based on this criterion, 9 participants were excluded. Applying the exact same criteria used to identify outliers in the previous experiments resulted in no further exclusions. The final sample included 69 participants (40 males), who ranged in age from 18 to 62 ( $M = 33.88$ ,  $SD = 11.61$ ).

**Materials and Procedures.** Six new texts, 3 in each condition, were used. The stories in the popular fiction condition were selected from an edited anthology of popular fiction (29). They were also chosen to represent a range of genres, including science fiction [*Space Jockey* by Robert Heinlein], mystery [*Too Many Have Lived* by Dashiell Hammett] and romance [*Lalla* by Rosamunde Pilcher]. Stories in the literary fiction condition were selected from a collection of the 2012 winners of the PEN/O. Henry Award for short literary fiction (30). They included *Corrie* by Alice Munroe, *Leak* by Sam Ruddick, and *Nothing Living Lives Alone* by Wendell Berry.

Originally posted 3 October 2013; revised 4 October 2013

After reading their assigned story, participants completed the RMET and the Author Recognition Test. All participants also completed the PANAS, answered a single item about current sadness, and reported their age, gender, ethnicity, and highest attained education.

#### Experiment 4

**Participants.** Participants ( $N = 106$ ) were recruited and compensated (\$2.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Since participants were recruited from the same pool as in the previous experiments, participants in this experiment (and Experiment 5) were asked whether they had participated in one of the prior experiments. Those who reported having done so were excluded ( $n = 13$ ). In this experiment (and Experiment 5), for the first time, participants also indicated their native language. One participant, who was not a native English speaker, was excluded. As in Experiment 3, texts were retyped, and applying the same criterion for excluding participants with low reading times (i.e.,  $< 30$ s) resulted in the exclusion of 14 participants. The exact same criteria as in the prior three experiments were used to delete outliers: one on the Author Recognition Test; five on error rates on the control trials of the Yoni task (described below); no participants on the RMET or on the critical Yoni trials. The final sample included 72 participants (20 males), who ranged in age from 18 to 59 ( $M = 34.26$ ,  $SD = 10.90$ ).

**Materials and Procedures.** Four of the texts used in Experiment 4 were the same as those used in Experiment 3. Two new texts, *Jane* by Mary Jane Rinehart (29, popular fiction) and *Uncle Rock* by Dagoberto Gilb (30, literary fiction), replaced *Lalla* (29) and *Leak* (30) from Experiment 3. After reading their assigned text, participants completed

Originally posted 3 October 2013; revised 4 October 2013

the RMET followed by the Yoni task, which assesses both cognitive and affective ToM (7). Finally, participants completed the Author Recognition Test and answered demographic questions including age, gender, and level of education.

The Yoni task includes first order (tracking one target's mental state) and second order (tracking two targets' mental states) cognitive (24 trials) and affective (24 trials) sections, as well as 14 control trials in which participants must only identify physical characteristics of images. For each trial, participants are shown four images surrounding a central character (usually named Yoni, but called John in this study). For the cognitive ToM trials, participants must identify, based on visual and linguistic cues, which of the four surrounding images the central character is thinking about or wants. For affective ToM trials, participants must make similar decisions about which image the central character likes, loves, dislikes, or does not love. Scores reflect proportion of correct responses. This task has been used in a wide range of populations to assess both components of ToM (7, 8, 31).

The RMET was administered in the same way as in Experiments 1 and 3.

### **Yoni Task Analysis**

Performance on the Yoni task was analyzed using a mixed analysis of covariance (ANCOVA) with type of trial (cognitive, affective) and trial difficulty (first order, second order) as within-participants factors. Experimental condition (literary fiction, popular fiction) and Author Recognition Test were between-participants factors, and performance on control trials was a covariate. Aside from the effects discussed in the main text, there was a main effect of type of trial [ $F(1,67) = 6.08, P = .01$ ], trial difficulty [ $F(1,67) = 6.43, P = .01$ ], and performance on control trials [ $\beta = 0.49, P < .001$ ]. There were significant

Originally posted 3 October 2013; revised 4 October 2013

interactions of trial difficulty and trial type [ $F(1,67) = 9.98, P = .002$ ] and of trial type and performance on control trials [ $F(1,67) = 5.35, P = .02$ ]. In addition, there was a three-way interaction of type of trial, trial difficulty, and performance on control trials [ $F(1,67) = 8.15, P = .005$ ].

### Experiment 5

**Participants.** Participants ( $N = 456$ ) were recruited and compensated (\$3.00) using Amazon.com's Mechanical Turk service and completed the experiment online using Qualtrics Software. Participants ( $n = 22$ ) who had completed one of the previous experiments were excluded. Participants who reported not being native English speakers were excluded ( $n = 9$ ). As in the previous two experiments page length was standardized and participants who spent less than 30 seconds per page of text were also excluded ( $n = 29$ ). The exact same criteria as in the prior four experiments were used to delete outliers: five on reading time; 24 on the Author Recognition Test guessing score; 9 on error rates on the control trials of the Yoni task; one on the RMET; 1 on the critical trials of the Yoni task. The final sample included 356 participants (182 males), who ranged in age from 18 to 75 ( $M = 34.42, SD = 11.59$ ).

**Materials and Procedures.** Five of the texts used in Experiment 5 were the same as those used in Experiment 4, but *The Vandercook* by Alice Mattinson (30) replaced *Nothing Alive Lives Alone* by Wendell Berry (30) in the literary fiction condition because it was shorter and so closer in length to the other texts. Participants completed the RMET followed by the Yoni task. Then, participants completed the PANAS and two items assessing current feelings of sadness and happiness on 7-point scales before completing the Author Recognition Test. Participants in the reading conditions then completed the

Originally posted 3 October 2013; revised 4 October 2013

Transportation scale and responded to two items assessing how much they enjoyed the text and how much they thought it was “an example of excellent literature” on 7-point scales. Finally, all participants answered demographic questions assessing their age, gender, ethnicity, and highest level of attained education.

### **Yoni Task Analysis**

As in Experiment 4, performance on the Yoni task was analyzed using a mixed ANCOVA with type of trial (cognitive, affective) and trial difficulty (first order, second order) as within-participants factors. Experimental condition (literary fiction, popular fiction, no-reading) and Author Recognition Test scores were entered as between-participants factors, and performance on control trials as covariate. Aside from the effects discussed in the main text, there was a main effect of difficulty of trial [ $F(1,351) = 28.99$ ,  $P < .001$ ] and a marginal effect of trial type [ $F(1,351) = 3.15$ ,  $P = .07$ ]. The effect of performance on control trials was significant [ $F(1,351) = 25.40$ ,  $P < .001$ ]. The interaction of trial difficulty and type of trial was significant [ $F(1,351) = 7.95$ ,  $P = .005$ ], as was the interaction of trial difficulty and performance on control trials [ $F(1,351) = 8.59$ ,  $P = .003$ ].