# Does Reading a Single Short Story of Literary Fiction Improve Social-Cognitive Skills? Testing the Priming Hypothesis

Jan Lenhart[1] and Tobias Richter[2]

[1] Department of Psychology, University of Bamberg

[2] Department of Psychology IV, University of Würzburg

**Manuscript accepted for publication in *Psychology of Aesthetics, Creativity, and the Arts* (2022)**

## Author Note

Jan Lenhart https://orcid.org/0000-0001-6235-7396

Tobias Richter https://orcid.org/0000-0002-0467-9044

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Jan Lenhart, Department of Psychology, University of Bamberg, Markusplatz 3, 96047 Bamberg, Germany. Phone: +(49)9518631994. E-Mail: jan.lenhart@uni-bamberg.de

# Does Reading a Single Short Story of Literary Fiction Improve Social-Cognitive Skills? Testing the Priming Hypothesis

## Abstract

Reading stories is a popular leisure activity in Western societies. Several current theories agree that reading might improve social cognition. Priming, in terms of an activation of content stored in long-term memory and facilitation of subsequent cognitive processing, has been proposed as a mechanism that leads to a temporary increase in social-cognitive task performance when reading a single story. In addition, this effect might be more pronounced given a rich prior reading experience. To test these hypotheses, we conducted two experiments in which participants either read a filler text and then a nonfiction text (nonfiction control condition), a narrative text and then a filler text (non-priming control condition), or a filler text and then a narrative text (priming condition). The participants completed a questionnaire on demographics and an author-recognition test. As dependent variables, two social-cognitive tasks on empathy and theory of mind were administered before and after reading the text stimuli (Experiment 1) or only after reading the text stimuli (Experiment 2). We found no significant differences between conditions on self-reported empathy or theory-of-mind performance in both experiments. Moreover, equivalence testing largely confirmed that the outcomes for the experimental and control conditions were statistically equivalent. Rich prior reading experience did not increase effects of narrative exposure. Accordingly, the results challenge the assumption that a brief exposure to narratives improves social-cognitive skills.

*Keywords:* narrative, literature, empathy, theory of mind, priming

## Introduction

Engaging in narratives in the form of books, films, or TV series is a common leisure-time activity for many children, adolescents, and adults in Western societies. Narratives refer to "actions and events which causally unfold in time" (Graesser et al., 1980, p. 283). Usually, they depict one or several characters whose intentions, goals, or plans encounter vicissitudes (Bruner, 1986) and consequently results in the character experiencing a variety of emotions (Oatley, 1999). These vicissitudes often deal with relationships among individuals, with love and conflict being the most common themes (Hogan, 2003). Accordingly, several theories assume that exposure to narratives in general or at least to so-called "literary" narratives might represent some kind of fictional training camp for social cognition, leading to readers' increased insight into their own and other's social-cognitive and emotional processes (e.g., Kidd & Castano, 2013; Koopman & Hakemulder, 2015; Mar, 2018; Oatley, 1999; for an overview, see Black et al., 2021).

Although these theories agree on the beneficial effect of engagement in narratives, the idea that reading a single short story or a single excerpt of a story might improve social-cognitive and emotional processes, such as perspective taking, has been challenged (e.g., Panero et al., 2016). The exposure to just one story might simply be too short to elicit any meaningful training effect on social-cognitive skills, which usually evolve during childhood and adolescence but remain quite stable during adulthood (e.g., Doris et al., 2022; Grühn et al., 2008; Quince et al., 2011). A more plausible alternative explanation for the effects of single stories on social-cognitive and emotional processes is that reading about characters' thoughts and feelings and their interactions with other individuals might activate corresponding content and processes

in long-term memory, resulting in a priming effect that temporarily improves social-cognitive task performance (Mumper & Gerrig, 2019; Panero et al., 2016). If this priming hypothesis holds, the effect of stories on social cognition should be short-lived and become weaker or even disappear if aspects of social cognition are assessed after a delay created by an intervening task (Roskos-Ewoldsen et al., 2009). The notion of priming has been used to explain media effects on a wide range of psychological variables (e.g., aggressiveness: cf. Bushman, 1998, social stereotypes: cf. Domke et al., 1999, political judgments: cf. Carpentier et al., 2008). However, to our knowledge, the priming hypothesis has not yet been tested systematically for the effects of stories on social-cognitive skills.

Accordingly, the goal of the present study was to examine the priming hypothesis in the context of an exposure to a single short story of literary fiction. In the following sections, we first briefly summarize different mechanisms that are proposed for explaining increases in social-cognitive task performance, focusing on the priming hypothesis. Then, we review empirical research that links social-cognitive task performance and story exposure, checking for compatibility with the priming hypothesis. Finally, we turn to the present study that comprises two experiments.

**How Narratives Might Affect Social-Cognitive Skills**

Although many theories agree on the beneficial effects of engagement in narratives, they propose different mechanisms by which narratives may exert these effects (e.g., Kidd & Castano, 2013; Koopman & Hakemulder, 2015; Mar, 2018; Oatley, 1999). So-called simulation theories assume that readers simulate characters' cognitions and emotions, running a multitude of simulation of minds when reading a story (Mar, 2018; Mar & Oatley, 2008; Oatley, 1999). These theories often refer to the

metaphor of a flight simulator, in which a (future) pilot trains to fly a plane in the real world (Oatley, 2016). Similar to the pilot's growing competence in flying the simulator that generalizes also to flying real planes, the simulation of minds is assumed to increase readers' understanding of the inner workings of fictional characters that generalizes also to real-life situations.

Other authors argue that simulating others' minds is not necessarily a prerequisite for understanding narratives and thus for explaining improvements in social cognition through narratives (Mumper & Gerrig, 2019). Instead, they argue that narratives offer many situations that are depicted from different points of view, that is, they provide information on divergences in knowledge, beliefs, and preferences and make those divergences salient, helping to improve perspective taking which is a core element of (cognitive) theory of mind (Mumper & Gerrig, 2019).

Additionally, many theories acknowledge that imitation learning by observing characters' actions and their consequences might be an important mechanism for acquiring social-cognitive knowledge and skills (Hakemulder, 2000; Mar, 2018; Mumper & Gerrig, 2019). Similar to real people, fictional characters might be suitable role models for acquiring social knowledge (Bandura, 1986). By reading a story, the reader might learn, for example, under which circumstances helping someone is required or how to express emotions in a socially appropriate manner.

The theoretical accounts described so far have proposed plausible mechanisms that might play a role in the effects of stories on social cognition and are supported by empirical research. However, whether they can explain the effects of short-term exposure to narrative fiction on social-cognitive skills is questionable. Much like pilots who will not learn to fly an airplane by spending just one hour in a flight simulator,

simulating a character's cognitions and emotions after reading just one story will not likely improve social-cognitive skills. A similar argument can be made with the ideas that stories influence social-cognitive skills because the characters serve as role models or facilitate perspective taking. The learning processes posited by these accounts are likely to take time and require frequent exposure, in particular because narratives are not designed to foster learning, and social-cognitive skills are relatively stable during adulthood. In Mar's (2018) words, given that "narratives appear to be a weak training context, any influence of stories on social cognitive processes and social knowledge seem likely to emerge only after prolonged, repeated, and frequent exposure" (p. 465).

If learning or training social-cognitive skills through stories takes time, how can the effects of short-term exposure to stories documented in the literature be explained? *Priming* in terms of activation of contents that are stored in long-term memory has been proposed as mechanisms to explain increases in social-cognitive task performance after reading a single short story or text excerpt (Mumper & Gerrig, 2019; Panero et al., 2016). In contrast to other potential mechanisms, the priming hypothesis does not assume that new content is actually learned or that processes are permanently improved. Instead, the activation of networks associated with social cognition and of already existing social-cognitive content stored in long-term memory is assumed to improve access to associated knowledge, skills, and processes and thereby influence the processing of subsequent information and increase performance in social-cognitive tasks temporarily (Förster & Liberman, 2007).

Empirically, priming effects through accessibility have been shown for different tasks such as faster response rates in decision tasks (Zwaan et al., 2002), improved recall in memory tasks (Shimamura & Squire, 1984), and activation of (negative)

stereotypes (Johnson et al., 2009) (for an overview, see Förster & Liberman, 2007). In addition, Wexler et al. (2016) found that administering a brief brain-training game (e.g., pattern recognition) designed to activate neural systems associated with executive function, immediately before math or reading games, increased speed and accuracy. Moreover, priming through media content, including the content of mediated stories, has been observed for diverse psychological outcomes (for an overview, see Roskos-Ewoldsen et al., 2009). Consequently, stories featuring social interactions among individuals and multiple individuals' thoughts and perceptions might serve as primes that activate social-cognitive content and processes and thereby increase task performance in social-cognitive tasks.

## Reviewing the Empirical Evidence for the Priming Hypothesis of Social-Cognitive Skills Through Stories

Since Kidd and Castano's (2013) *Science* paper, which created a stir at that time, an increasing number of studies have examined the effects of narratives on social-cognitive task performance. Kidd and Castano (2013) conducted five experiments in which they examined the immediate effects of reading short texts of literary fiction, popular fiction, nonfiction, and no-reading on theory of mind. The subsequent studies typically compared reading a narrative story versus reading nonfiction (e.g., Bal & Veltkamp, 2013; Black & Barnes, 2015; Chlebuch et al., 2020), or they compared literary fiction to popular fiction or both (e.g., Kidd & Castano, 2019; Panero et al., 2016).

On a meta-analytic level, the first comparison yields a small effect ($g$ = 0.15/ 0.16) in favor of narratives (Dodell-Feder & Tamir, 2018). The second comparison has not yet been summarized systematically. The research findings seem to be decidedly

mixed for the comparison between literary and popular texts (Kidd & Castano, 2013, 2019; Kidd et al., 2016; Panero et al., 2016; Pino & Mazza, 2016; Samur et al., 2018; van Kuijk et al., 2018). To further complicate the matter, some findings have shown that the effects of text stimuli might depend on individual differences such as reading experience (Kidd & Castano, 2019), trait openness to experience (Djikic et al., 2013), and transportation (Bal & Veltkamp, 2013). Kidd and Castano (2019, Experiment 1), for example, found that individuals with more reading experience showed higher theory-of-mind performance after reading a literary text than those reading a popular text, whereas no differences were observed for individuals with lower reading experience.

In most studies (e.g., Black & Barnes, 2015; Kidd & Castano, 2013, 2019; Panero et al., 2016; Samur et al., 2018; van Kuijk et al., 2018), participants read a single short story or short excerpt from a book and then immediately took the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001), which requires the participants to gauge the mental or emotional state of a person based on the eyes of the person's face and is assumed to measure theory of mind. This study design precluded any determination of whether increases in RMET performance reflect real improvements or a priming effect. These explanations, namely training and priming effects, do not necessarily exclude each other and might even be complementary.

Although the results of many studies are compatible or even better compatible with the priming hypothesis, research that systematically compares the priming hypothesis to those hypotheses that assume real improvements in social-cognitive skills is still lacking. Conceptually, priming effects depend on two conditions, the intensity of the stimulus and the time that has passed between the processing of the stimulus (e.g.,

reading the story) and the assessment of the outcome (e.g., a measure of social cognition) (Roskos-Ewoldsen et al., 2009). The first condition is not unique to priming effects, but it also pertains to other types of media effects (including the training of social skills), whereas the second condition can facilitate distinguishing between the priming hypothesis and alternative accounts. For media and social priming, priming effects usually dissipate within minutes. According to the mental model account of priming proposed by Roskos-Ewoldsen et al. (2009), they may be expected to occur as long as the mental model or situation model constructed during processing the story is active in memory. Thus, creating a condition with an intervening cognitive task between reading the story and the posttest, such as reading a text on another topic, should basically eliminate or at least reduce a priming effect.

**Purpose of the Present Study**

The present study was conducted to examine the hypothesis that reading a short narrative about human interaction can activate associated content in long-term memory and thus temporarily influence subsequent cognitive processing and increase performance in social-cognitive tasks. Although this priming hypothesis is compatible with the results of many previous studies (e.g., Black & Barnes, 2015; Kidd & Castano, 2013; van Kuijk et al., 2018), it has not yet been examined systematically.

To systematically examine the priming hypothesis, we conducted two experiments in which participants either read a nonfiction filler text and a nonfiction text (nonfiction condition), a narrative text and a nonfiction filler text (non-priming condition), or a nonfiction filler text and a narrative text (priming condition). To allow for a priming effect in the narrative priming condition, the participants completed the social-cognitive tasks directly after reading the second text. Given that the average

effect of narratives on social-cognitive skills tends to be small ($g$ = 0.15/0.16; Dodell-Feder & Tamir, 2018) and literary narratives are sometimes found to have a stronger impact than popular narratives (e.g., Kidd & Castano, 2013; Kidd et al., 2016; van Kuijk et al., 2018), we selected a literary short story from the studies conducted by Kidd and Castano (2013) for the narrative conditions to maximize any potential effect ($d$ = 0.51 literary fiction vs. nonfiction in Kidd & Castano, 2013, Study 1, equal sample sizes for each experimental group assumed). In the first experiment, we used a pre-post design so that we could examine changes in social-cognitive skills and increase the precision of the design. In the second experiment, the pretest was eliminated to exclude the possibility that pretesting of social-cognitive skills might cause a testing effect irrespective of the text that is read afterwards.

In line with the priming hypothesis, we expected to find a stronger increase in social-cognitive skills in the narrative priming condition compared to the nonfiction condition and the non-priming condition. In addition, we expected no difference between the non-priming condition and the nonfiction condition. However, an alternative prediction could be that exposure to a short narrative might actually improve (and not just prime) social-cognitive skills. In this case, a different pattern of results would be expected to emerge. The narrative conditions should not differ in their effects on social-cognitive skills, but both should lead to higher scores in social-cognitive performance than the nonfiction condition.

Finally, some tentative evidence indicates that individuals' reading experience might increase effects of narratives on social-cognitive skills (Kidd & Castano, 2019). We therefore explored whether individuals leisure reading moderates the effect of

reading a narrative text (compared to a nonfiction text) on social-cognitive skills, expecting that higher amounts of leisure reading increases the effect.

## Experiment 1

### Methods

Detailed information concerning the materials and the procedure as well as the data and analysis script are available on OSF (https://osf.io/upvqy/?view_only=f32bcffe794045078b001cb935620f2b).

#### *Sample*

Participants were recruited online by advertising in social media and the University of (anonymized)'s online systems for study participation. Participants received 7.50 Euro or study credits for their participation. The credit option was only available for students in specific study courses at the university. All participants consented to participate and confirmed by selecting several boxes that they had read and understood the study information (e.g., procedure, data protection, voluntary participation) before they started the study. The project was approved by the university's ethics committee.

Assuming an effect size of $d = 0.507$ for the difference between literary fiction and nonfiction (Kidd & Castano, 2013, Study 1, equal sample sizes for the experimental conditions assumed), an a-priori power analysis with PANGEA (Westfall, 2016) indicated that a sample of at least 42 participants per condition (or 126 participants in total) would be needed to detect a significant difference in pre- to posttest changes between groups (3 x 2 mixed design; $\alpha = .05$; $1\text{-}\beta = .80$). In addition, following the small-telescopes approach (Simonsohn, 2015), setting the effect size of interest as the effect size that the study conducted by Kidd and Castano (2013; Study 1) would have

had 33% power to detect ($d_{33\%}$ = 0.332), a power analysis with PANGEA indicated that a sample of at least 96 participants per condition (or 288 participants in total) were needed.

Out of the 348 subjects that started the study, 224 remained in the sample after applying several exclusion criteria (see the section *Data Preparation and Statistical Analyses*). The final sample consisted of 175 females and 49 males who were mostly students (82%) and had a mean age of 25.47 years (*SD* = 11.17). Accordingly, most participants reported a higher education entrance qualification (72%) or a university degree (22%) as their highest educational level. All participants had at least good communication skills in German, with the vast majority being native speakers (97%). For a full description of the sample see Table 1.

### *Instruments*

**Demographics.** Participants provided information on their age (in years), their gender (male, female, diverse), their education (1 = *no graduation* to 5 = *university degree;* 6 = *other type of educational qualification*), their student status (yes/no), and their German language proficiency level (1 = *basic communication skills* to 5 = *native speaker*).

**Leisure Reading.** A German Author Recognition Test (ART; Grolig et al., 2020; checklist B) was used to assess narrative leisure reading. Participants are required to indicate whether or not they know an author by selecting a box. The ART consists of 26 highbrow and 24 popular authors and contains 25 foils. For each correctly chosen author, 1 point is awarded, resulting in scores ranging from 0 to 26 for highbrow literature, 0 to 24 for popular literature, and 0 to 50 for total leisure reading. For comparability between scales, the scores were dived by the maximum per scale to

obtain a value between 0 and 1, with higher scores indicating more leisure reading. Foil items were not included in the analyses but were used to check for individuals who might distort their responses.[1] With none of the participants of the original sample marking more than five (20%) of the foils, none of the responses were conspicuous. In addition, to check for the possibility that the ART was inadvertently skipped, participants were required to select at least one of the boxes or to indicate that they knew none of the names. The latter option was used by only one participant in the original sample. Internal consistency was excellent ($\omega_{total}$ = .91).

**Theory of Mind.** A German version (Bölte, 2005) of the well-known and widely used Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001) was used to assess theory of mind. The RMET consists of 36 items that require participants to gauge the mental or emotional state of a person based on photos that depict only the eyes of that person. Each photo had four options with one correct answer and three distractors. The RMET score ranges between 0 and 36, with higher scores reflecting better theory of mind. Internal consistency was low for the pre- and the posttest ($\omega_{total}$ = .51 and .61).

---

[1] Grolig et al. (2020) calculated a corrected hit rat by subtracting the proportion of false hits (foils) from the proportion of correct hits. In the present study, we deemed the number of correct hits to be the better approximation of reading experience. However, we calculated all analyses in Experiment 1 and 2 that included the ART a second time using the corrected hit rate. The results did not change. Thus, we report only the analyses using the uncorrected hit rate in the manuscript. All analyses are available in the R scripts on osf.

**Empathy.** The Saarbrücker Persönlichkeitsfragebogen (IRI-S D, V 7.0; Paulus, 2019), a German adaption of the Interpersonal Reactivity Index (IRI; Davis, 1980), was used to assess trait empathy. The IRI-S D consists of 16 items that are answered on a 5-point Likert scale (1 = *never* to 5 = *always*), with four items assessing Personal Distress, Perspective Taking, Empathic Concern, and Fantasy respectively. The latter three subscales are combined to form the Empathy scale (Paulus, 2012), with mean scores ranging from 1 to 5 and higher scores reflecting higher empathy. Internal consistency was acceptable at the pre- and the posttest ($\omega_{total}$ = .79 and .85).

### *Text Stimuli and Control Questions*

A German translation of the short story, *The Runner* (DeLillo, 2012; 2,174 words), which was used in previous studies as literary text (Kidd & Castano, 2013), was chosen as the narrative stimulus. In *The Runner,* a runner observes an abduction in a park and talks about this incident with other people. We used as nonfiction text *Der Siegeszug der Teufelsknolle* (The triumph of the devil's bulb; Bayerischer Rundfunk, 2019; 1,457 words), an informative text about the history of the potato. As a nonfiction filler text, we used the text *Wie Pflanzen den Tau aufsaugen* (How plants soak up the dew; Schlichting, 2020; 846 words) in which the interaction between leaves and water absorption is described.

To ensure that participants had read the texts, each text was followed by three easy comprehension questions. The questions were presented in the single-choice format with three options, with one correct guess representing the chance level.

### *Design and Procedure*

Experiment 1 was based on a 2 x 3 mixed design, with measurement point (pre- vs. posttest) representing the within-subjects factor and experimental condition

(nonfiction condition vs. non-priming condition vs. priming condition) representing the between-subjects factor. The experiment, which took approximately 45 min to complete, was conducted with the online tool SoSci Survey (www.soscisurvey.de), which is a free survey tool for non-commercial research. Before the study started, participants received information concerning the study (e.g., procedure, data protection, voluntary participation) and were required to consent to participate by selecting several boxes. At the beginning, participants were asked demographic questions. Then they were pretested on theory of mind (RMET) and empathy (IRI-S D), which was followed by the assessment of leisure reading (ART). Then participants were randomly assigned to one of the three conditions. They either read the filler text followed by the nonfiction text (nonfiction condition), the narrative text followed by the filler text (non-priming condition), or the filler text followed by the narrative text (priming condition). After each text the participants answered three easy single-choice questions to control for text comprehension. Then posttest comprising the RMET and the IRI-S D were administered. At the end of the study, participants received further information concerning the study goal and were instructed how to proceed to receive credits or payment.

### *Data Preparation and Statistical Analyses*

**Exclusion Criteria.** We used several exclusion criteria to add more control to the experiment because online participation cannot be equally well controlled as experiments in the lab. Starting from 348 subjects who had begun the study, we excluded participants based on the following sequence of exclusion criteria: they (1) did not complete the study ($n$ = 41), (2) did not answer the easy control questions for the nonfiction or the narrative text above chance level (i.e., at least 2 out of 3; $n$ = 13), (3)

read the text stimuli faster than 650 words per minute, indicating scanning (Carver, 1992) and being much faster than normal silent reading speed (Brysbaert, 2019) ($n$ = 62), and (4) had values on any of the measures used in the analyses (RMET, IRI-S D, ART) that were univariate outliers (> 3 median absolute deviations; Leys et al., 2013; $n$ = 8). After applying these exclusion criteria, the final sample consisted of 224 subjects, with similar group sizes in the experimental conditions (priming: $n$ = 73; non-priming: $n$ = 79; non-fiction: $n$ = 72).

**Statistical Analyses.** All statistical analyses were conducted with R (Version 4.2.0). We used the R package *Routliers* (Version 0.0.0.3; Delacre & Klein, 2019) for detecting outliers and the package *psych* (Version: 2.2.5; Revelle, 2022) for calculating internal consistencies of the scales (*omega total*). We employed mixed-effects modeling (*lmer*) provided in the R packages *lme4* (Version: 1.1-29; Bates et al., 2015) to test the hypotheses and *lmerTest* (Version 3.1-3; Kuznetsova et al., 2017) to obtain $p$ values for the coefficients. The R package *emmeans* (Version 1.7.4-1; Lenth, 2022) was used to compare individual groups by calculating pairwise interaction contrasts on the mixed-effects models and to calculate effect sizes (Cohen's $d$; Cohen, 1988) for pairwise comparisons. We used the R package *Desctools* (Version 0.99.45; Signorell et al., 2022) for calculating effect sizes for the contingency tables and the packages *sjPlot* (Version 2.8.10; Lüdecke, 2021) and *ggplot2* (Version 3.3.6; Wickham, 2016) to create tables and figures for the mixed-effects models. Categorical variables were effect-coded and continuous predictor variables were mean-centered for the mixed-effects analyses. In addition, following guidelines provided in Lakens et al. (2018), we conducted tests of equivalence between the conditions. Using the small-telescopes approach (Simonsohn, 2015) to determine the bounds of equivalence (i.e., the effect size of interest), we tested

against an effect size of $d = 0.332$, which represents the mean effect size that could have been detected with a power of .33 ($d_{33\%}$) in Kidd and Castano (2013, Study 1). Inspection of the data indicated that all variables with the exception of the ART scores were normally distributed. The ART scores were positively skewed, which were corrected by applying a square root transformation. The alpha level was set to $p < .05$ (two-tailed).

## Results

### *Preliminary Analyses*

Descriptive statistics for the whole sample and the experimental conditions are displayed in Table 1. There was no indication of floor or ceiling effects in any of the dependent variables. Additionally, we found no statistically significant differences between the groups in any of the demographic variables (gender: $\chi^2 = 2.96$, $p = .233$, Cramer's $V = .11$, 95%-CI [.00, .23]; student status: $\chi^2 = 0.95$, $p = .630$, Cramer's $V = .07$, 95%-CI [.00, .17]; education: $\chi^2 = 13.25$, $p = .090$, Cramer's $V = .17$, 95%-CI [.00, .22]; age: $F(2, 221) = 0.30$, $p = .738$, $\omega^2 < .01$, 95%-CI [.00, .00]; language proficiency: $F(2, 221) = 0.75$, $p = .476$, $\omega^2 < .01$, 95%-CI [.00, .00]), the scores of the author recognition test, $F(2, 221) = 2.23$, $p = .110$, $\omega^2 = .01$, 95%-CI [.00, .05], or the dependent variables at the pretest, theory of mind: $F(2, 221) = 0.79$, $p = .456$, $\omega^2 < .01$, 95%-CI [.00, .00]; empathy: $F(2, 221) = 0.96$, $p = .384$, $\omega^2 < .01$, 95%-CI [.00, .00].

Empathy scores ($r = .92$, $p < .001$, 95%-CI [.90, .94]) and theory-of-mind scores ($r = .66$, $p < .001$, 95%-CI [.58, .73]) showed a high stability of individual differences from pre- to posttest. Both socio-cognitive measures were only weakly correlated at pretest ($r = .18$, $p = .006$, 95%-CI [.05, .31]) and posttest ($r = .19$, $p = .005$, 95%-CI [.06, .31]). In addition, empathy scores ($r = -.06$, $p = .395$, 95%-CI [-.19, .07]) and

theory-of-mind scores ($r$ = -.08, $p$ = .248, 95%-CI [-.21, .05]) were not significantly correlated with leisure reading at the pretest.

***Effects on Theory of Mind***

As can be seen from Table 2 and Figure 1, we found a significant effect of time, with theory-of-mind scores increasing from pretest to posttest. None of the other main or interaction effects were statistically significant. Pairwise interaction contrasts showed no difference between the priming and non-priming conditions, *estimate* = -0.78, $t$(221) = -1.59, $p$ = .113, $d$ = -0.21, 95%-CI [-0.48, 0.05], and the priming and nonfiction conditions, *estimate* = -0.75, $t$(221) = -1.50, $p$ = .135, $d$ = -0.21, 95%-CI [-0.48, 0.06]. We also found no significant difference between the non-priming and the nonfiction conditions, *estimate* = 0.03, $t$(221) = 0.06, $p$ = .956, $d$ = 0.01, 95%-CI [-0.26, 0.27].

[Insert Figure 1 here]

The tests of equivalence ($d$ = 0.332) were significant for the comparison of the non-priming and the nonfiction conditions (lower bound: $p$ = .006; upper bound: $p$ = .008) but not for the comparison between priming and non-priming conditions (lower bound: $p$ = .188; upper bound: $p$ < .001) and the priming and nonfiction conditions (lower bound: $p$ = .179; upper bound: $p$ < .001), indicating that only the non-priming and nonfiction conditions were statistically equivalent.

***Effects on Empathy***

As can be seen from Table 2 and Figure 2, we found no significant main or interaction effects. Pairwise interaction contrasts showed no difference between the priming and non-priming condition, *estimate* = -0.00, $t$(221) = -0.01, $p$ = .991, $d$ = -0.00,

95%-CI [-0.13, 0.13], and the priming and nonfiction condition, *estimate* = 0.01, $t(221) = 0.32$, $p = .746$, $d = 0.02$, 95%-CI [-0.11, 0.15]. We also found no difference between the non-priming and the nonfiction conditions, *estimate* = 0.01, $t(221) = 0.34$, $p = .733$, $d = 0.02$, 95%-CI [-0.11, 0.15].

[Insert Figure 2 here]

The tests of equivalence ($d = 0.332$) were significant for the comparison of the non-priming and the nonfiction conditions (lower bound: $p < .001$; upper bound: $p < .001$), the priming and the nonfiction conditions (lower bound: $p < .001$; upper bound: $p < .001$) and the priming and the non-priming conditions (lower bound: $p < .001$; upper bound: $p < .001$), indicating that the conditions were statistically equivalent.

As suggested by a reviewer, we also conducted the analyses for empathy separately for each of the four subscales of the IRI-S D. As in the analysis using the empathy score, there were no statistically significant interaction effects between time and condition in any of the subscales of the IRI-SD (see Table A1 in the Appendix).

***Examining Leisure Reading as a Moderator***

**Theory of Mind.** After including the ART scores as a moderator, the positive effect of time on RMET scores remained significant (see Table 2). However, the analysis revealed a significant three-way interaction between time, condition, and ART scores. Pairwise interaction contrasts at the mean and ± 1 *SD* of the ART scores showed that there were no significant differences between conditions at –1 *SD* ($p > .254$) and at the mean ($p > .175$). However, in contrast to the hypothesis, the priming conditions resulted in lower RMET scores than the non-priming condition at +1 *SD*, *estimate*

= -1.67, $t(221) = -2.52$, $p = .012$, $d = -0.46$, 95%-CI [-0.83, -0.10], and the nonfiction condition, *estimate* = -2.23, $t(221) = -3.16$, $p = .002$, $d = -0.62$, 95%-CI [-1.01, -0.23], whereas no difference was found between the non-priming and the nonfiction conditions ($p = .462$).

**Empathy.** After including the ART scores as a moderator, the effect of time and any of its interactions with condition and ART scores were not significant (see Table 2). Pairwise interaction contrasts at the mean and ±1 *SD* of the ART scores showed no significant differences between conditions at –1 *SD* ($p > .274$), at the mean ($p > .815$), and at +1 *SD* ($p > .357$).

As suggested by a reviewer, we also conducted the analyses for empathy separately for each of the four subscales of the IRI-S D. As in the analysis using the empathy score, we found no statistically significant interaction effects between time, condition, and leisure reading in the Fantasy, Perspective Taking, and Personal Distress subscales of the IRI-SD (see Table 2 in the Appendix). In the Emotional Concern subscale, a three-way-interaction emerged (see Table A2 in the Appendix). However, in contrast to the hypothesis, pairwise interaction contrasts indicated that at low rates of leisure reading, participants in the non-priming condition led to more self-reported emotional concern than those in the non-fiction condition.

## Discussion

In Experiment 1, we used a pre-post design to investigate the effects of reading a literary narrative and the hypothesis that any increases in social-cognitive task performance could be caused by an activation of social-cognitive content stored in long-term memory. To examine this priming hypothesis, we compared a narrative priming condition, a narrative non-priming condition, and a nonfiction condition. We found that

performance in the theory-of-mind task increased irrespective of the condition. However, the priming condition was neither significantly different from the other conditions nor statistically equivalent to the other conditions. Thus, the comparisons including the priming condition have low informational value. Self-reported empathy was not affected by any of the conditions and showed no change from pretest to posttest. Finally, individual differences in leisure reading had no influence on the effect of narratives on self-reported empathy. However, leisure reading moderated the effect on theory-of-mind performance, with the priming condition resulting in less gains than the other conditions for frequent readers.

These results are not consistent with theories that propose beneficial learning effects of short-term exposures to (literary) narratives on social-cognitive skills (e.g., Kidd & Castano, 2013, 2019). They are also inconsistent with the priming hypothesis that assumes activation of social-cognitive skills results in an increase in social-cognitive task performance (e.g., Panero et al., 2016). Given that we found a positive change in theory-of-mind performance across conditions, our pre-post-design might have led to these outcomes. More precisely, the pretests of our social-cognitive measures might have influenced the post-test measures (e.g., in the form of memory or expectation effects), perhaps covering any priming effects of the narratives. Alternatively, RMET scores might have increased across conditions because of a testing effect. To control for these possibilities, we conducted a second experiment in which we dropped the pretests of the social-cognitive measures.

## Experiment 2

### Methods

Experiment 2 used the same instruments, text stimuli, and experimental procedure as the first experiment. However, the method had no pretest of social-cognitive measures, and we changed the order of the social-cognitive tasks in the posttest, starting with the self-report measure of empathy (IRI-S D) instead of the theory-of-mind performance task (RMET) because potential carry-over effects between measures should be smaller in this order.

Detailed information concerning the materials and the procedure as well as the data and analysis script are available on OSF (https://osf.io/upvqy/?view_only=f32bcffe794045078b001cb935620f2b).

***Sample***

The sample consisted of participants that were recruited online by advertising in the University of (anonymized)'s online system for study participation. Participants received study credits for their participation. All participants consented and confirmed their participation by selecting several boxes that they had read and understood the study information (e.g., procedure, data protection, voluntary participation) before they started the study. The project was approved by the university's ethics committee.

Assuming an effect size of $d = 0.507$ for the difference between literary fiction and nonfiction (Kidd & Castano, 2013, Study 1, equal sample sizes assumed), an a-priori power analysis with PANGEA (Westfall, 2016) indicated that at least 62 participants per condition or 186 participants in total would be needed to detect a significant difference between groups (3-group between-subjects design; $\alpha = .05$; $1\text{-}\beta = .80$). In addition, following the small-telescopes approach (Simonsohn, 2015), setting the effect size of interest as the effect size of the study conducted by Kidd and Castano (2013; Study 1) would have had 33% power to detect ($d_{33\%} = 0.332$), a power analysis

with PANGEA indicated that a sample of at least 143 participants per condition (or 429 participants in total) were needed.

Out of the 562 subjects that started the study, 408 remained in the sample after applying several exclusion criteria (see the section *Data Preparation and Statistical Analyses*). The final sample consisted of 354 females and 54 males who were mostly students (> 99%) and had a mean age of 21.14 years (*SD* = 4.18). Accordingly, most participants reported a higher education entrance qualification (89%) or a university degree (9%) as their highest educational level. All participants had at least good communication skills in German, with the vast majority being native speakers (98%). See Table 3 for a full description of the sample.

Based on the suggestion of a reviewer, we also added post-hoc a no-reading control group in which participants only completed the social-cognitive measures but read no text before. This extra control group consisted of 158 participants. Applying the same exclusion criteria as in the original sample, 147 participants remained in the post-hoc no-reading control group.

### *Instruments, Text Stimuli, and Control Questions*

Instruments, text stimuli, and control questions were identical to those used in Experiment 1. Internal consistency was low for the RMET ($\omega_{total}$ = .47), acceptable for the IRI-S D ($\omega_{total}$ = .79), and good for the ART ($\omega_{total}$ = .89).

### *Design and Procedure*

Experiment 2 was based on a between-subjects design (nonfiction condition vs. non-priming condition vs. priming condition) and took approximately 30 min to complete on soscisurvey.de. The procedure was similar to that of Experiment 1. However, we excluded a pretest of the dependent variables and changed the order of the

social-cognitive tasks in the posttest, starting with the self-report measure (IRI-S D) instead of the performance task (RMET) to reduce potential carry-over effects between social-cognitive measures. An additional no-reading control group that was not part of the randomized experiment was added post-hoc during the review process based on a reviewer's suggestion.

### *Data Preparation and Statistical Analyses*

**Exclusion Criteria.** We used the same exclusion criteria as in the first experiment. Starting from 562 subjects who had begun the study, we excluded participants based on the following sequence of exclusion criteria: they (1) did not complete the study ($n$ = 39), (2) did not answer the easy control questions for the nonfiction or the narrative text above chance (at least 2 out of 3; $n$ = 22), (3) read the text stimuli faster than 650 words per minute ($n$ = 75), and (4) had values on any of the measures used in the analyses (RMET, IRI-S D, ART) that were univariate outliers (> 3 median absolute deviations; Leys et al., 2013; $n$ = 18). After applying these exclusion criteria, the final sample consisted of 408 subjects, with similar group sizes in the experimental conditions (priming: $n$ = 138; non-priming: $n$ = 136; non-fiction: $n$ = 134). In the post-hoc no-reading control group, 147 participants of the starting sample ($N$ = 158) remained because six did not complete the study and five had values on any of the measures used in the analyses that were univariate outliers (using the cut-off criteria of the original sample).

**Statistical Analyses.** Statistical analyses were again conducted with R. The statistical procedure and the analyses were similar to those in Experiment 1. However, because of the pretest elimination, we employed linear modeling (*lm*) to test the hypotheses. Inspection of the data indicated that all variables with the exception of the

ART scores were normally distributed. The ART scores were positively skewed, which were corrected by applying a square root transformation. The alpha level was set to $p < .05$ (two-tailed).

## Results

### *Preliminary Analyses*

Descriptive statistics for the whole sample and the experimental conditions are displayed in Table 3. As can be seen, there was no indication of floor or ceiling effects in any of the dependent variables. Additionally, we found no statistically significant differences between the groups in any of the demographic variables (gender: $\chi^2 = 0.78$, $p = .681$, Cramer's $V = .04$, 95%-CI [.00, .12]; student status: $\chi^2 = 2.05$, $p = .328$, Cramer's $V = .07$, 95%-CI [.00, .16]; education: $\chi^2 = 7.88$, $p = .470$, Cramer's $V = .10$, 95%-CI [.00, .13]; age: $F(2, 405) = 0.12$, $p = .884$, $\omega^2 < .01$, 95%-CI [.00, .00]; language proficiency: $F(2, 405) = 1.63$, $p = .197$, $\omega^2 < .01$, 95%-CI [.00, .02]), or the scores of the author recognition test, $F(2, 405) = 0.52$, $p = .595$, $\omega^2 < .01$, 95%-CI [.00, .00] .

Self-reported empathy and theory-of-mind scores were weakly correlated ($r = .12$, $p = .013$, 95%-CI [.03, .22]). Both showed weak positive relations to leisure reading (theory of mind: $r = .12$, $p = .016$, 95%-CI [.02, .21]; empathy: $r = .09$, $p = .065$, 95%-CI [-.01, .19]).

### *Effects on Theory of Mind*

As can be seen from Table 4 and Figure 3, there was no effect of experimental condition on theory-of-mind performance. Pairwise contrasts revealed a nonsignificant difference between the priming and non-priming conditions, *estimate* = -0.01, $t(405) = -0.02$, $p = .986$, $d = -0.00$, 95%-CI [-0.24, 0.24], and between the priming and nonfiction conditions, *estimate* = -0.02, $t(405) = -0.06$, $p = .956$, $d = -0.01$, 95%-CI [-0.25, 0.23].

We also found no significant difference between the non-priming and the nonfiction conditions, *estimate* = -0.01, $t(405) = -0.04$, $p = .970$, $d = -0.00$, 95%-CI [-0.24, 0.24].

[Insert Figure 3 here]

The tests of equivalence were significant for the comparison of the non-priming and the nonfiction conditions (lower bound: $p = .004$; upper bound: $p = .003$), the priming and the nonfiction conditions (lower bound: $p = .004$; upper bound: $p = .003$), and the priming and the non-priming conditions (lower bound: $p = .003$; upper bound: $p = .003$), indicating that the conditions were statistically equivalent.

***Effects on Empathy***

As can be seen from Table 4 and Figure 4, there was no effect of experimental condition on self-reported empathy. Pairwise contrasts revealed a nonsignificant difference between the priming and non-priming conditions, *estimate* = 0.01, $t(405) = 0.13$, $p = .900$, $d = 0.02$, 95%-CI [-0.22; 0.25], and between the priming and nonfiction conditions, *estimate* = 0.04, $t(405) = 0.80$, $p = .424$, $d = 0.10$, 95%-CI [-0.14; 0.34]. We also found no significant difference between the non-priming and the nonfiction conditions, *estimate* = 0.03, $t(405) = 0.67$, $p = .502$, $d = 0.08$, 95%-CI [-0.16; 0.32].

[Insert Figure 4 here]

The tests of equivalence were significant for the comparison of the non-priming and the nonfiction conditions (lower bound: $p < .001$; upper bound: $p = .020$), the priming and the nonfiction conditions (lower bound: $p < .001$; upper bound: $p = .027$),

and the priming and the non-priming conditions (lower bound: $p = .002$; upper bound: $p = .005$), indicating that the conditions were statistically equivalent.

As suggested by a reviewer, we also conducted the analyses of empathy separately for each of the four subscales of the IRI-S D. We also found no significant effect of experimental condition in any of the subscales of the IRI-SD (see Table A3 in the Appendix).

### *Examining Leisure Reading as a Moderator*

**Theory of Mind.** After including the ART scores as a moderator, the main effect of condition, the main effect of leisure reading, and their interaction were nonsignificant (see Table 4). Pairwise contrasts showed no significant differences between conditions at –1 *SD* ($p > .516$), at the mean ($p > .922$), and at +1 *SD* ($p > .474$) of the ART scores.

**Empathy.** After including the ART scores as a moderator, the main effect of condition, the main effect of leisure reading, and their interaction were nonsignificant (see Table 4). Pairwise contrasts showed no significant differences between conditions at –1 *SD* ($p > .310$), at the mean ($p > .380$), and at +1 *SD* ($p > .510$) of the ART scores.

As suggested by a reviewer, we also conducted the analyses on empathy separately for each of the four subscales of the IRI-S D. We found no statistically significant interaction between experimental condition and leisure reading in any of the subscales of the IRI-SD (see Table A4 in the Appendix).

### *Post-Hoc Analyses: Comparison to a Post-Hoc No-Reading Control Group*

To further disambiguate the results, we also compared the experimental conditions to a no-reading control group that received only the measures but no stimulus text. The no-reading control group was tested post hoc and was therefore not part of the

randomized experiment. Similar to the other conditions, the participants were in their early 20s, predominantly female, had a high educational level, consisted of students only, and German was the native language for most of the participants (see Table 3).

**Theory of Mind.** Similar to the original analysis without the no-reading control group, no significant effect of condition on theory-of-mind performance was found (see Table A5). Pairwise contrasts revealed a nonsignificant difference between the no-reading condition and the priming, *estimate* = -0.27, $t(551) = -0.71$, $p = .479$, $d = -0.08$, 95%-CI [-0.32; 0.15], the non-priming, *estimate* = -0.26, $t(551) = -0.69$, $p = .492$, $d = -0.08$, 95%-CI [-0.32; 0.15], and the nonfiction, *estimate* = -0.24, $t(551) = -0.65$, $p = .518$, $d = -0.08$, 95%-CI [-0.31; 0.16] conditions. In addition, the tests of equivalence were significant for the comparison of the no-reading condition with the priming (lower bound: $p = .018$; upper bound: $p < .001$), non-priming (lower bound: $p = .018$; upper bound: $p < .001$), and nonfiction (lower bound: $p = .017$; upper bound: $p < .001$) conditions, indicating that the original experimental conditions were statistically equivalent to the no-reading control group.

**Empathy.** Similar to the original analysis without the no-reading control group, we found no significant effect of condition on theory-of-mind performance (see Table A5). Pairwise contrasts revealed no significant differences between the no-reading condition and the priming, *estimate* = -0.01, $t(551) = -0.28$, $p = .781$, $d = -0.03$, 95%-CI [-0.27; 0.20], the non-priming, *estimate* = -0.02, $t(551) = -0.41$, $p = .683$, $d = -0.05$, 95%-CI [-0.28; 0.19], and the nonfiction, *estimate* = -0.05, $t(551) = -1.11$, $p = .267$, $d = -0.13$, 95%-CI [-0.37; 0.10] conditions. In addition, the tests of equivalence were significant for the comparison of the no-reading condition with the priming (lower bound: $p = .006$; upper bound: $p = .001$), non-priming (lower bound: $p = .009$; upper

bound: $p < .001$), and nonfiction (lower bound: $p = .048$; upper bound: $p < .001$) conditions, indicating that the original experimental conditions were statistically equivalent to the no-reading control group.

**Leisure Reading as Moderator.** Similar to the original analysis without the no-reading control group, no significant interaction effect was found between experimental condition and participants' leisure reading on theory-of-mind performance and self-reported empathy (see Table A5).

### Discussion

In Experiment 2, we used a posttest-only design to investigate the effects of reading a literary narrative and to test the hypothesis that any increases in social-cognitive task performance could be caused by an activation of social-cognitive content stored in long-term memory. To examine this priming hypothesis, we compared a narrative priming condition, a narrative non-priming condition, a nonfiction condition, and a post-hoc no-reading control group. We found no influence of condition on self-reported empathy and theory-of-mind performance. Finally, individual differences in leisure reading played no moderating role in the effect of narratives on self-reported empathy or theory-of-mind performance.

## General Discussion

In the present study, we examined whether reading a short story increases self-reported empathy and performance in a theory-of-mind task on a short-term basis. In particular, our aim was to examine the priming hypothesis, which assumes that increases in social-cognitive self-reports or performance following an exposure to a single short story are due to an activation of social-cognitive content in long-term memory (Mumper & Gerrig, 2019; Panero et al., 2016). To test this hypothesis, we

compared a priming group, a non-priming group, and a nonfiction control condition. In Experiment 2, we added post hoc a no-reading control group. We also explored whether prior reading experience moderates any effect of narratives on social-cognitive skills, which was reported by Kidd and Castano (2019).

In contrast to our hypotheses, we found no significant differences between conditions in Experiment 1 and 2. Given the positive change in theory-of-mind performance across conditions in Experiment 1, we ran an additional experiment to rule out the possibility that the null results found for the experimental groups were an artefact of the pre-post-design used in Experiment 1. More precisely, the pretests of our social-cognitive measures might have created specific expectations or their responses might have been remembered by participants, perhaps covering any priming effects of the narratives. Alternatively, RMET scores might have increased across conditions because of a testing effect. To control for these possibilities, we conducted a second experiment in which we dropped the pretests of the social-cognitive measures. Again, Experiment 2 revealed no differences between conditions on social-cognitive task performance. Therefore, we consider the idea that the inclusion of a pretest in Experiment 1 masked effects of narrative exposure on social-cognitive skills to be implausible. Moreover, with the exception of small effect sizes (around $d = 0.20$) to the disadvantage of the priming condition in Experiment 1, effect sizes of group comparisons were close to zero across both experiments. We also employed equivalence testing (Lakens et al, 2018) to test whether the outcomes for the experimental groups were equivalent, which we found for all theoretically relevant comparisons. These results are inconsistent with theories that propose beneficial learning effects of short-term exposures to (literary) narratives on social-cognitive skills (e.g., Kidd & Castano,

2013, 2019). They also are not consistent with the priming hypothesis (e.g., Mumper & Gerrig, 2019; Panero et al., 2016). Nonetheless, our findings add to a decidedly mixed research of experimental studies (e.g., Black & Barnes, 2015; Kidd & Castano, 2013, 2019; Panero et al., 2016; Samur et al., 2018; van Kuijk et al., 2018), with an overall small effect size and large heterogeneity of effects (Dodell-Feder & Tamir, 2018).

The heterogeneity of effect sizes could have occurred because of moderator variables such as transportation into the story (Mar, 2018) or participants' prior reading experience (Kidd & Castano, 2019). In both experiments, we examined whether prior reading experiences moderated effects of narratives on social-cognitive skills, which was found by Kidd and Castano (2019). Across both experiments, we found no evidence that more leisure reading increased effects of narratives. In contrast to the assumption and Kidd and Castano's (2019) finding, leisure reading moderated the effect on theory-of-mind performance in Experiment 1, with the priming condition resulting in less gains than the other conditions for frequent readers. We could not replicate this finding in Experiment 2, and we have no sensible explanation why participants with richer prior reading experience should be less influenced by a priming effect. Therefore, this finding should be treated with caution and should not be over-interpreted. It might have been due to chance.

Of course, we cannot conclude from our findings that social cognition cannot be influenced through priming or that it cannot be improved. Previous research suggests that social cognition can be primed, for example, social stereotypes (e.g., Johnson et al., 2009). However, a short exposure to stories might not be (reliably) suitable to either prime or improve social-cognitive skills such as theory-of-mind performance or self-reported trait empathy in adults. Moreover, although there seems to be no general

beneficial effect of a short exposure to narratives on social-cognitive skills, potential moderator variables, such as narrative transportation into the stories, could produce these effects (Bal & Veltkamp, 2013; Mar, 2018; Schwerin & Lenhart, 2022). Finally, given the findings of correlational studies, which typically report small but significant correlations between leisure reading and social-cognitive skills (e.g., Lenhart et al., 2020; Mar et al., 2006; for a meta-analysis see Mumper & Gerrig, 2017), it seems likely that frequent, cumulative engagement with (literary) fiction might indeed train social-cognitive skills in the long run, as suggested by some authors (e.g., Koopman & Hakemulder, 2015; Mar, 2018; Oatley, 1999).

**Limitations and Directions for Future Research**

The present study has some limitations that need to be discussed. A first limitation is that we used only two measures to assess social-cognitive skills, namely a self-report measure of trait empathy (IRI-S D) and a performance measure of theory of mind (RMET). Although the RMET is widely used in research (e.g., Dodell-Feder & Tamir, 2018), it has been repeatedly shown to have low internal consistency in the German and the English version (e.g., Mar et al., 2006; Meyer & Shean, 2006; Schwerin & Lenhart, 2022), indicating that more reliable measures are needed to assess social-cognitive performance. A second limitation is that we used two measures of social-cognitive skills, which required us to determine an order of testing. This testing sequence in turn might have impacted the results. For example, it is possible that any priming effects might have faded out after the measurement of the first dependent variable. However, as the order of the dependent variables was reversed in Experiment 2, but the results were similar in Experiment 1 and 2, this seems to be an unlikely explanation. A third limitation is that we used only a single narrative text. Although this

literary story, which had been included in the seminal study conducted by Kidd and Castano (2013), was selected based on the assumption that literary fiction should be even better suited to enhance social-cognitive skills (Kidd & Castano, 2013), the extent that our results can be generalized to studies that use other literary stories is unclear. A fourth limitation is that our sample consisted predominantly of highly-educated female participants. Although the meta-analysis conducted by Dodell-Feder and Tamir (2018) on narrative effects on social-cognitive skills found that the percentage of female participants and the sample type (students vs. mechanical Turks) were not significant moderators, the generalizability of our results to other populations is unclear. Finally, when we turn to theories that assume real (and therefore longer-lasting) improvements of social-cognitive skills through narratives (e.g., Mar, 2018; Oatley, 1999), the operationalization of the present study and of many other studies (e.g., Kidd & Castano, 2013, 2019; Panero et al., 2016; Samur et al., 2018) that have provided only a single short exposure to a narrative stimulus are inconsistent with the theoretical assumption that frequent exposure to narratives cumulatively builds social knowledge and hones social processes. Thus, the results of these type of short-term experimental studies cannot be used to determine whether or not this assumption might be true. To answer this question, well-designed longitudinal correlational studies and longitudinal high-intensity intervention studies are needed.

## References

Bal, P. M., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS ONE, 8*(1), e55341. https://doi.org/10.1371/journal.pone.0055341

Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Prentice-Hall

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*, 241–251. https://doi.org/10.1111/1469-7610.00715

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bayerischer Rundfunk. (2019, 16. August). *Der Siegeszug der Teufelsknolle* [The triumph of the devil's bulb]. https://www.br.de/wissen/kartoffel-geschichte-europa100.html

Black, J. E., & Barnes, J. L. (2015). The effects of reading material on social and non-social cognition. *Poetics*, *52,* 32–43. https://doi.org/10.1016/j.poetic.2015.07.001

Black, J. E., Barnes, J. L., Oatley, K., Tamir, D. I., Dodell-Feder, D., Richter, T., & Mar, R.A. (2021). Stories and their role in social cognition. In D. Kuiken & A.M. Jacobs (Eds.), *Handbook of empirical literary studies* (pp. 241–260). De Gruyter. https://doi.org/10.1515/9783110645958-010

Bölte, S. (2005). *Reading mind in the eyes test: Erwachsenenversion* [Reading the mind in the eyes test: Adult version]. https://docs.autismresearchcentre.com/tests/EyesTest_German.pdf

Bruner, J. (1986). *Actual minds, possible worlds*. Harvard University Press.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language, 109*, Article 104047. https://doi.org/10.1016/j.jml.2019.104047

Bushman, B. J. (1998). Priming effects of media violence on the accessibility of aggressive constructs in memory. *Personality and Social Psychology Bulletin, 24*(5), 537–545. https://doi.org/10.1177/0146167298245009

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading, 36*(2), 84–95.

Chlebuch, N., Goldstein, T. R., & Weisberg, D. S. (2020). Fact or fiction? Clarifying the relationship between reading and the improvement of social skills. *Scientific Study of Literature, 10*(2), 167–192. https://doi.org/10.1075/ssol.20007.chl

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Delacre, M., & Klein, O. (2019). *Routliers: Robust outliers detection.* R package version 0.0.0.3. https://CRAN.R-project.org/package=Routliers

DeLillo, D. (2012). *Der Engel Esmeralda: Neun Erzählungen* (translated by F. Heibert) [The angel Esmeralda]. Kiepenheuer & Witsch.

Dillman Carpentier, F. R., Roskos-Ewoldsen, D. R., & Roskos-Ewoldsen, B. B. (2008). A test of the network models of political priming. *Media Psychology, 11*(2), 186–206. https://doi.org/10.1080/15213260801994220

Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Reading other minds: Effects of literature on empathy. *Scientific Study of Literature*, *3*(1), 28–47. https://doi.org/10.1075/ssol.3.1.06dji

Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. *Journal of Experimental Psychology: General, 147*(11), 1713–1727. https://doi.org/10.1037/xge0000395

Domke, D. McKoy, K., & Torres, M. (1999). News media, racial perceptions, and political cognition. *Communication Research, 26*, 570–607. https://doi.org/10.1177/009365099026005003

Dorris, L., Young, D., Byrne, K., & Hoyle, R. (2022). Cognitive empathy across the lifespan. *Developmental Medicine & Child Neurology*. Advance online publication. https://doi.org/10.1111/dmcn.15263

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Förster, J., & Liberman, N. (2007). Knowledge activation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology. Handbook of basic principles* (2nd ed., p. 201–231). The Guilford Press.

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980). Advanced outlines, familiarity, and text genre on retention of prose. *Journal of*

*Experimental Education, 48*(4), 281–290. https://doi.org/10.1080/00220973.1980.11011745

Grolig, L., Tiffin-Richards, S. P., & Schroeder, S. (2020). Print exposure across the reading life span. *Reading and Writing*: *An Interdisciplinary Journal*, *33*(6), 1423–1441. https://doi.org/10.1007/s11145-019-10014-3

Grühn, D., Rebucal, K., Diehl, M., Lumley, M., & Labouvie-Vief, G. (2008). Empathy across the adult lifespan: Longitudinal and experience-sampling findings. *Emotion, 8*(6), 753–765. https://doi.org/10.1037/a0014123

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

Hogan, P.C. (2003). *The mind and its stories*. Cambridge University Press.

Johnson, J. D., Olivo, N., Gibson, N., Reed, W., & Ashburn-Nardo, L. (2009). Priming media stereotypes reduces support for social welfare policies: The mediating role of empathy. *Personality and Social Psychology Bulletin, 35*(4), 463–476. https://doi.org/10.1177/0146167208329856

Kidd, D., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*, 377–380. https://doi.org/10.1126/science.1239918

Kidd, D., & Castano, E. (2019). Reading literary fiction and theory of mind: Three preregistered replications and extensions of Kidd and Castano (2013). *Social Psychology and Personality Science, 10*(4), 522–531. https://doi.org/10.1177/1948550618775410

Kidd, D., Ongis, M., & Castano, E. (2016). On literary fiction and its effects on theory of mind. *Scientific Study of Literature*, *6*(1), 42–58. https://doi.org/10.1075/ssol.6.1.04kid

Koopman, E., & Hakemulder, J. F. (2015). Effects of literature on empathy and self-reflection: A theoretical-empirical framework. *Journal of Literary Theory, 9*(1), 79–111.

Kuznetsova A., Brockhoff P. B., Christensen R. H. B (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lenhart, J., Dangel, J., & Richter, T. (2020). The relationship between lifetime book reading and empathy in adolescents: Examining transportability as a moderator. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000341

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means.* R package version 1.7.4-1. https://CRAN.R-project.org/package=emmeans

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, *32*(1), 5. https://doi.org/10.5334/irsp.289

Lüdecke D (2021). sjPlot: Data visualization for statistics in social science. R package version 2.8.10. https://CRAN.R-project.org/package=sjPlot.

Mar, R. A. (2018). Evaluating whether stories can promote social cognition: Introducing the Social Processes and Content Entrained by Narrative (SPaCEN) framework. *Discourse Processes*, *55*(5-6), 454–479. https://doi.org/10.1080/0163853X.2018.1448209

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science, 3*, 173–192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Mar, R. A., Oatley, K., Hirsh, J. dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality, 40*, 694–712. https://doi.org/10.1016/j.jrp.2005.08.002

Meyer, J., & Shean, G. (2006). Social-cognitive functioning and schizotypal characteristics. *The Journal of Psychology, 140*(3), 199–207. https://doi.org/10.3200/JRLP.140.3.199-207

Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 109–120. http://dx.doi.org/10.1037/aca0000089

Mumper, M. L., & Gerrig, R. J. (2019). How does leisure reading affect social cognitive abilities? *Poetics Today, 40*(3), 453–473. https://doi.org/10.1215/03335372-7558080

Oatley, K. (1999). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Review of General Psychology, 3*, 101–117. https://doi.org/10.1037/1089-2680.3.2.101

Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences, 20*(8), 618–628. https://doi.org/10.1016/j.tics.2016.06.002

Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *Journal of Personality and Social Psychology*, *111*(5), e46-e54. http://dx.doi.org/10.1037/pspa0000064

Paulus, C. (2019). *Saarbrücker Persönlichkeitsfragebogen (IRI-S D). Based on the Interpersonal Reactivity Index (IRI). V 7.0*. Universität des Saarlandes. http://bildungswissenschaften.uni-saarland.de/personal/paulus/homepage/files/SPF-IRI-_V7.0.pdf

Pino, M.C., & Mazza, M. (2016). The use of "literary fiction" to promote mentalizing ability. *PLoS ONE 11*(8): e0160254. https://doi.org/10.1371/journal.pone.0160254

Quince, T. A., Parker, R. A., Wood, D. F., & Benson, J. A. (2011). Stability of empathy among undergraduate medical students: A longitudinal study at one UK medical school. *BMC Medical Education, 11*(1), 1–9.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R- project.org/

Revelle, W. (2022) *psych: Procedures for personality and psychological research*. Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version = 2.2.5.

Roskos-Ewoldsen, D. R., & Roskos-Ewoldsen, B. (2009). Media priming: An updated synthesis. In J. Bryant & M. B. Oliver (Eds.), *Media effects* (pp. 90–109). Routledge.

Samur, D., Tops, M., & Koole, S. L. (2018). Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013). *Cognition and Emotion, 32*(1), 130–144. https://doi.org/10.1080/02699931.2017.1279591

Schlichting, H. J. (2020, 1. June). *Wie Pflanzen den Tau aufsaugen* [How plants soak up the dew]. https://www.spektrum.de/wissen/wie-pflanzen-den-tau-aufsaugen/1721294

Schwerin, J., & Lenhart, J. (2022). The effects of literariness on social-cognitive skills: Examining narrative engagement, transportation, and identification as moderators. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000514

Shimamura, A. P., & Squire, L. R. (1984). Paired-associate learning and priming effects in amnesia: A neuropsychological study. *Journal of Experimental Psychology: General, 113*(4), 556–570. https://doi.org/10.1037/0096-3445.113.4.556

Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arachchige, C., Arppe, A., Baddeley, A., Barton, K., Bolker, B., Borchers, H. W., Caeiro, F., Champely, S., Chessel, D., Chhay, L., Cooper, N., Cummins, C., Dewey, M., Doran, H.

C.,… Zeileis, A. (2022). *DescTools: Tools for descriptive statistics*. R package version 0.99.45. https://cran.r-project.org/package=DescTools

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569. https://doi.org/10.1177/0956797614567341

van Kuijk, I., Verkoeijen, P., Dijkstra, K., & Zwaan, R. (2018). The effect of reading a short passage of literary fiction on theory of mind: A replication of Kidd and Castano (2013). *Collabra: Psychology, 4*(1). https://doi.org/10.1525/collabra.117

Westfall, J. (2016). PANGEA: *Power analysis for general anova designs. Working paper: Manuscript last updated October 11, 2016*. http://jakewestfall.org/publications/pangea.pdf

Wexler, B. E., Iseli, M., Leon, S., Zaggle, W., Rush, C., Goodman, A., Esat Emal, A., & Bo, E. (2016). Cognitive priming and cognitive training: Immediate and far transfer to academic skills in children. *Scientific Reports, 6,* 32859. https://doi.org/10.1038/srep32859

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

**Table 1**

*Descriptive Statistics of the Sample in Study 1*

| | **Final Sample (*N* = 224)** | | **Priming (*n* = 73)** | | **Non-Priming (*n* = 79)** | | **Nonfiction (*n* = 72)** | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Gender (female) | 175 (78%) | | 61 (84%) | | 57 (72%) | | 57 (79%) | |
| Student (yes) | 183 (82%) | | 57 (78%) | | 66 (84%) | | 60 (83%) | |
| Education | | | | | | | | |
| *Middle Track* | 4 (2%) | | 0 (0%) | | 1 (1%) | | 3 (4%) | |
| *Vocational Training* | 6 (3%) | | 1 (1%) | | 2 (3%) | | 3 (4%) | |
| *High Track* | 161 (72%) | | 50 (68%) | | 55 (70%) | | 56 (78%) | |
| *University Degree* | 49 (22%) | | 19 (26%) | | 21 (27%) | | 9 (13%) | |
| *Other* | 4 (2%) | | 3 (4%) | | 0 (0%) | | 1 (1%) | |
| Language Proficiency | 4.95 | 0.30 | 4.99 | 0.12 | 4.94 | 0.40 | 4.93 | 0.31 |
| Age (years) | 25.47 | 11.17 | 26.29 | 11.31 | 24.91 | 10.41 | 25.26 | 11.91 |
| ART (total) | 0.27 | 0.16 | 0.31 | 0.18 | 0.26 | 0.15 | 0.26 | 0.14 |
| ART (highbrow) | 0.31 | 0.18 | 0.34 | 0.20 | 0.30 | 0.17 | 0.28 | 0.16 |
| ART (popular) | 0.24 | 0.17 | 0.28 | 0.18 | 0.22 | 0.16 | 0.23 | 0.15 |
| Theory of Mind (t1) | 25.55 | 3.64 | 25.79 | 3.64 | 25.73 | 3.81 | 25.11 | 3.45 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Theory of Mind (t2) | 26.15 | 3.66 | 25.88 | 4.13 | 26.59 | 3.46 | 25.94 | 3.36 |
| Empathy (t1) | 3.72 | 0.42 | 3.77 | 0.39 | 3.68 | 0.42 | 3.71 | 0.44 |
| Empathy (t2) | 3.70 | 0.45 | 3.75 | 0.42 | 3.66 | 0.44 | 3.69 | 0.50 |

*Note.* The educational system in Germany comprises three high-school tracks in secondary school: Haupt-/Mittelschule (low track), Realschule (middle track), Gymnasium (high track), with only the latter track qualifying for university entrance. ART = Author Recognition Test (Grolig et al., 2020; checklist B). ART scores range from 0 to 1. Theory of Mind scores range from 0 to 36. Empathy scores range from 1 to 5.

**Table 2**

*Effects of the Experimental Conditions and of Leisure Reading on Theory of Mind and Empathy in Study 1*

| | **Theory of Mind** | | | | | | **Empathy** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* |
| Intercept | 25.84 | 116.15 | **<.001** | 25.86 | 116.42 | **<.001** | 3.71 | 129.36 | **<.001** | 3.71 | 128.61 | **<.001** |
| Condition (priming) | -0.01 | -0.02 | .982 | 0.09 | 0.27 | .785 | 0.05 | 1.26 | .209 | 0.05 | 1.25 | .214 |
| Condition (non-priming) | 0.32 | 1.04 | .300 | 0.36 | 1.15 | .251 | -0.04 | -1.07 | .286 | -0.05 | -1.24 | .217 |
| Time (posttest) | 0.30 | 2.94 | **.004** | 0.34 | 3.41 | **.001** | -0.01 | -1.81 | .072 | -0.01 | -1.93 | .055 |
| Condition (priming) x Time | -0.25 | -1.78 | .077 | -0.23 | -1.60 | .111 | 0.00 | 0.18 | .855 | -0.00 | -0.04 | .969 |
| Condition (non-priming) x Time | 0.13 | 0.96 | .340 | 0.10 | 0.74 | .462 | 0.00 | 0.21 | .837 | 0.00 | 0.22 | .823 |
| ART | | | | -1.85 | -1.28 | .203 | | | | -0.15 | -0.79 | .431 |
| Condition (priming) x ART | | | | -1.39 | -0.72 | .471 | | | | 0.25 | 0.99 | .322 |
| Condition (non-priming) x ART | | | | 5.40 | 2.64 | **.009** | | | | -0.59 | -2.22 | **.027** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Time x ART | | 0.54 | 0.83 | .407 | | 0.64 | .521 |
| Condition (priming) x Time x ART | | -2.70 | -3.10 | **.002** | | 1.07 | .286 |
| Condition (non-priming) x Time x ART | | 0.53 | 0.57 | .568 | | -1.36 | .175 |
| **Random Effects** | | | | | | | |
| Residual Variance | 4.55 | | 4.40 | | 0.02 | | 0.02 |
| Intercept Variance | 8.80 | | 8.60 | | 0.18 | | 0.17 |
| ICC | .66 | | .66 | | .92 | | .92 |
| N | 224 | | 224 | | 224 | | 224 |
| Observations | 448 | | 448 | | 448 | | 448 |
| Marginal $R^2$ / Conditional $R^2$ | .01 / .66 | | .05 / .68 | | .01 / .92 | | .03 / .92 |

*Note*. Categorical variables were effect coded (Time: pretest = -1; posttest = 1; Condition: nonfiction = -1; priming = 1; non-priming = 1) and continuous variables were centered. ART = Author Recognition Test (Grolig et al., 2020; checklist B).

**Table 3**

*Descriptive Statistics of the Sample in Study 2*

| | **Final Sample (*N* = 408)[a]** | | **Priming (*n* = 138)** | | **Non-Priming (*n* = 136)** | | **Nonfiction (*n* = 134)** | | **No-reading (*n* = 147)[b]** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Gender (female) | 354 (87%) | | 119 (86%) | | 116 (85%) | | 119 (89%) | | 130 (95%) | |
| Student (yes) | 407 (>99%) | | 138 (100%) | | 136 (100%) | | 133 (99%) | | 147 (100%) | |
| Education | | | | | | | | | | |
| *Middle Track* | 1 (<1%) | | 0 (0%) | | 1 (<1%) | | 0 (0%) | | 0 (0%) | |
| *Vocational Training* | 2 (<1%) | | 0 (0%) | | 1 (<1%) | | 1 (<1%) | | 3 (2%) | |
| *High Track* | 365 (89%) | | 124 (90%) | | 120 (88%) | | 121 (90%) | | 139 (95%) | |
| *University Degree* | 38 (9%) | | 14 (10%) | | 14 (10%) | | 10 (7%) | | 5 (3%) | |
| *Other* | 2 (<1%) | | 0 (0%) | | 0 (0%) | | 2 (1%) | | 0 (0%) | |
| Language Proficiency | 4.95 | 0.33 | 4.98 | 0.19 | 4.96 | 0.31 | 4.91 | 0.43 | 4.96 | 0.26 |
| Age (years) | 21.14 | 4.18 | 21.06 | 3.37 | 21.09 | 4.04 | 21.29 | 5.02 | 20.71 | 2.88 |
| ART (total) | 0.25 | 0.13 | 0.24 | 0.14 | 0.25 | 0.14 | 0.26 | 0.13 | 0.22 | 0.13 |
| ART (highbrow) | 0.27 | 0.14 | 0.26 | 0.14 | 0.27 | 0.14 | 0.28 | 0.14 | 0.25 | 0.13 |
| ART (popular) | 0.23 | 0.16 | 0.23 | 0.15 | 0.23 | 0.17 | 0.23 | 0.15 | 0.19 | 0.15 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Theory of Mind | 26.46 | 3.10 | 26.45 | 3.03 | 26.46 | 3.29 | 26.47 | 3.01 | 26.71 | 3.28 |
| Empathy | 3.72 | 0.41 | 3.74 | 0.40 | 3.73 | 0.41 | 3.70 | 0.43 | 3.75 | 0.37 |

*Note.* The educational system in Germany comprises three high-school tracks in secondary school: Haupt-/Mittelschule (low track), Realschule (middle track), Gymnasium (high track), with only the latter track qualifying for university entrance. ART = Author Recognition Test (Grolig et al., 2020; checklist B). ART scores range from 0 to 1. Theory of Mind scores range from 0 to 36. Empathy scores range from 1 to 5.

[a] The final sample refers to the sample of the experimental design after applying the exclusion criteria. The no-reading control group that was added post hoc during the review process is not included.

[b] The no-reading control group was added post hoc during the review process and was thus not part of the randomized experimental design.

**Table 4**

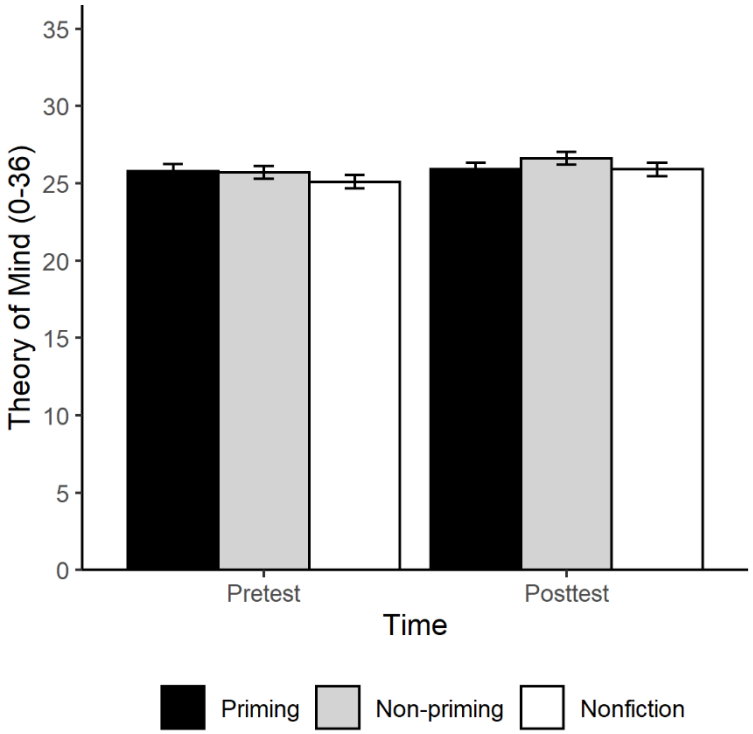
*Effects of the Experimental Conditions and of Leisure Reading on Theory of Mind and Empathy in Study 2*

| | Theory of Mind | | | | | | Empathy | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* |
| Intercept | 26.46 | 171.74 | **<.001** | 26.46 | 172.33 | **<.001** | 3.72 | 181.27 | **<.001** | 3.72 | 181.28 | **<.001** |
| Condition (priming) | -0.01 | -0.04 | .966 | 0.02 | 0.09 | .926 | 0.02 | 0.54 | .592 | 0.02 | 0.59 | .556 |
| Condition (non-priming) | -0.00 | -0.01 | .991 | -0.00 | -0.01 | .989 | 0.01 | 0.32 | .752 | 0.01 | 0.35 | .729 |
| ART | | | | 2.83 | 2.42 | **.016** | | | | 0.30 | 1.90 | .058 |
| Condition (priming) x ART | | | | 1.20 | 0.74 | .462 | | | | -0.17 | -0.78 | .438 |
| Condition (non-priming) x ART | | | | -1.53 | -0.93 | .350 | | | | 0.12 | 0.55 | .582 |
| Observations | | 408 | | | 408 | | | 408 | | | 408 | |
| $R^2$ / $R^2$ adjusted | | .00 / -.01 | | | .02 / .00 | | | .00 / -.00 | | | .01 / -.00 | |

*Note*. Categorical variables were effect coded (Condition: nonfiction = -1; priming = 1; non-priming = 1) and continuous variables were centered. ART = Author Recognition Test (Grolig et al., 2020; checklist B).

**Figure 1.**

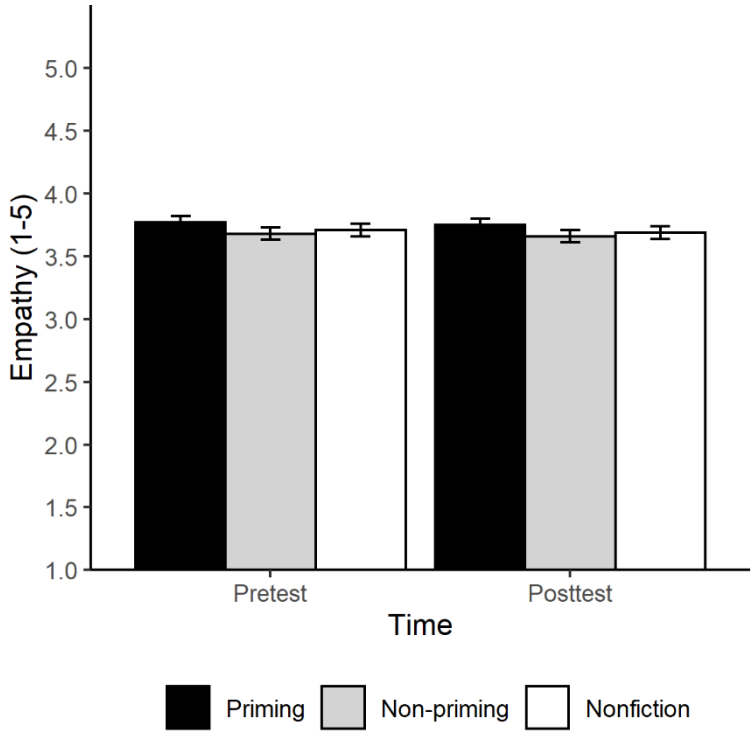
*Effects of the Experimental Conditions on Theory of Mind in Study 1*




*Note.* Error bars represent standard errors.

**Figure 2.**

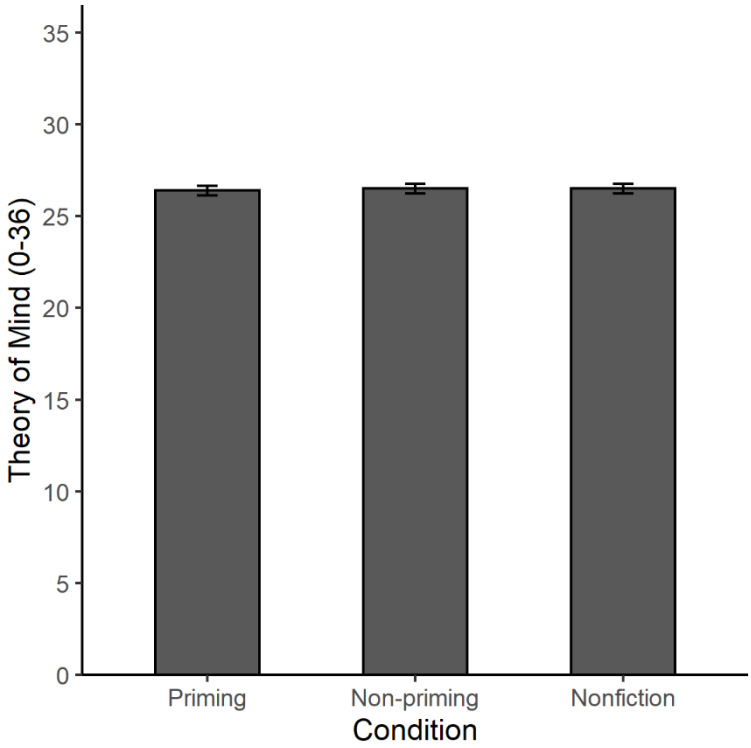
*Effects of the Experimental Conditions on Empathy in Study 1*




*Note.* Error bars represent standard errors.

**Figure 3.**

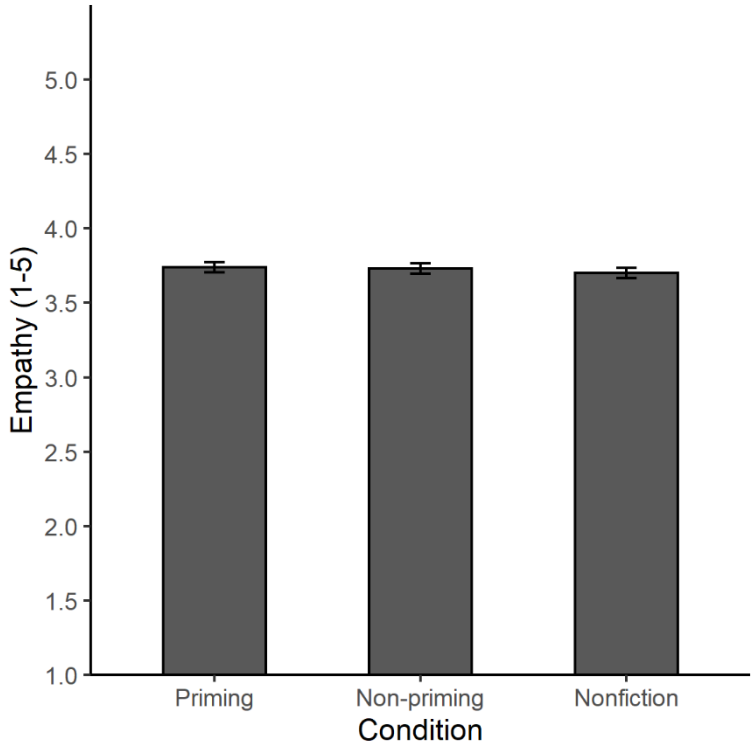
*Effects of the Experimental Conditions on Theory of Mind in Study 2*




*Note.* Error bars represent standard errors.

**Figure 4.**

*Effects of the Experimental Conditions on Empathy in Study 2*




*Note.* Error bars represent standard errors.

## Appendix

## Table A1

*Effects of the Experimental Conditions on the Subscales of the IRI-S D in Study 1*

| | **Emotional Concern** | | | **Perspective Taking** | | | **Fantasy** | | | **Personal Distress** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* |
| Intercept | 3.88 | 103.42 | **<.001** | 3.61 | 92.28 | **<.001** | 3.64 | 79.56 | **<.001** | 2.82 | 55.29 | **<.001** |
| Condition (priming) | 0.01 | 0.14 | .887 | 0.09 | 1.60 | .110 | 0.06 | 0.88 | .379 | -0.04 | -0.54 | .587 |
| Condition (non-priming) | 0.04 | 0.70 | .482 | 0.00 | 0.08 | .935 | -0.17 | -2.66 | **.008** | -0.06 | -0.87 | .383 |
| Time (posttest) | -0.01 | -1.28 | .200 | -0.01 | -0.63 | .533 | -0.02 | -1.45 | .149 | -0.05 | -4.43 | **<.001** |
| Condition (priming) x Time | -0.01 | -0.52 | .602 | 0.01 | 1.03 | .306 | -0.00 | -0.31 | .760 | 0.00 | 0.20 | .842 |
| Condition (non-priming) x Time | 0.01 | 0.92 | .358 | -0.01 | -0.77 | .444 | 0.01 | 0.41 | .679 | -0.00 | -0.26 | .797 |
| **Random Effects** | | | | | | | | | | | | |
| Residual Variance | | 0.03 | | | 0.05 | | | 0.05 | | | 0.06 | |

| | | | | |
|---|---|---|---|---|
| Intercept Variance | 0.30 | 0.32 | 0.44 | 0.55 |
| ICC | .92 | .87 | .89 | .90 |
| N | 224 | 224 | 224 | 224 |
| Observations | 448 | 448 | 448 | 448 |
| Marginal $R^2$ / Conditional $R^2$ | .00 / .92 | .02 / .87 | .03 / .90 | .01 / .90 |

*Note*. Categorical variables were effect coded (Condition: nonfiction = -1; priming = 1; non-priming = 1).

**Table A2**

*Effects of the Experimental Conditions and of Leisure Reading on the Subscales of the IRI-S D in Study 1*

| | **Emotional Concern** | | | **Perspective Taking** | | | **Fantasy** | | | **Personal Distress** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *Statistic* | *p* | *Estimates* | *Statistic* | *p* | *Estimates* | *Statistic* | *p* | *Estimates* | *Statistic* | *p* |
| Intercept | 3.88 | 103.01 | **<.001** | 3.61 | 92.24 | **<.001** | 3.63 | 79.96 | **<.001** | 2.81 | 54.41 | **<.001** |
| Condition (priming) | 0.03 | 0.49 | .622 | 0.06 | 1.15 | .252 | 0.06 | 0.97 | .331 | -0.03 | -0.42 | .673 |
| Condition (non-priming) | 0.03 | 0.53 | .600 | 0.01 | 0.18 | .859 | -0.19 | -2.95 | **.004** | -0.07 | -0.90 | .367 |
| Time (posttest) | -0.01 | -1.34 | .183 | -0.01 | -0.65 | .514 | -0.02 | -1.58 | .114 | -0.05 | -4.43 | **<.001** |
| ART | -0.58 | -2.38 | **.018** | 0.62 | 2.43 | **.016** | -0.48 | -1.63 | .106 | -0.36 | -1.05 | .293 |
| Condition (priming) x Time | -0.01 | -0.54 | .587 | 0.01 | 0.83 | .407 | -0.01 | -0.47 | .640 | 0.00 | 0.14 | .888 |
| Condition (non-priming) x Time | 0.01 | 0.72 | .474 | -0.01 | -0.62 | .538 | 0.01 | 0.45 | .654 | -0.00 | -0.12 | .908 |
| Condition (priming) x ART | 0.02 | 0.05 | .958 | 0.22 | 0.66 | .509 | 0.50 | 1.28 | .203 | 0.21 | 0.47 | .641 |
| Condition (non-priming) x ART | -0.10 | -0.28 | .777 | -0.48 | -1.32 | .188 | -1.20 | -2.86 | **.005** | -0.11 | -0.22 | .823 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time x ART | -0.02 | -0.43 | .670 | 0.07 | 1.10 | .273 | 0.02 | 0.30 | .762 | 0.02 | 0.27 | .789 |
| Condition (priming) x Time x ART | 0.04 | 0.64 | .523 | 0.02 | 0.18 | .861 | 0.11 | 1.13 | .261 | 0.03 | 0.26 | .795 |
| Condition (non-priming) x Time x ART | -0.17 | -2.40 | **.017** | 0.05 | 0.50 | .614 | -0.10 | -1.02 | .308 | 0.10 | 0.92 | .361 |
| **Random Effects** | | | | | | | | | | | | |
| Residual Variance | | 0.03 | | | 0.05 | | | 0.05 | | | 0.06 | |
| Intercept Variance | | 0.30 | | | 0.31 | | | 0.42 | | | 0.56 | |
| ICC | | 0.92 | | | 0.87 | | | 0.89 | | | 0.90 | |
| N | | 224 | | | 224 | | | 224 | | | 224 | |
| Observations | | 448 | | | 448 | | | 448 | | | 448 | |
| Marginal $R^2$ / Conditional $R^2$ | | .03 / .92 | | | .05 / .87 | | | .07 / .90 | | | .02 / .90 | |

*Note*. Categorical variables were effect coded (Condition: nonfiction = -1; priming = 1; non-priming = 1) and continuous variables were centered. ART = Author Recognition Test (Grolig et al., 2020; checklist B).

**Table A3**

*Effects of the Experimental Conditions on the Subscales of the IRI-S D in Study 2*

| | **Emotional Concern** | | | **Perspective Taking** | | | **Fantasy** | | | **Personal Distress** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* |
| Intercept | 3.90 | 141.12 | **<.001** | 3.65 | 125.92 | **<.001** | 3.62 | 113.61 | **<.001** | 2.87 | 90.04 | **<.001** |
| Condition (priming) | 0.02 | 0.42 | .677 | 0.05 | 1.32 | .186 | -0.02 | -0.53 | .598 | 0.01 | 0.15 | .881 |
| Condition (non-priming) | 0.01 | 0.38 | .702 | 0.01 | 0.27 | .788 | 0.00 | 0.04 | .971 | -0.02 | -0.38 | .703 |
| Observations | 408 | | | 408 | | | 408 | | | 408 | | |
| $R^2$ / $R^2$ adjusted | .00 / -.00 | | | .01 / .00 | | | .00 / -.00 | | | .00 / -.01 | | |

*Note*. Categorical variables were effect coded (Condition: nonfiction = -1; priming = 1; non-priming = 1).

**Table A4**

*Effects of the Experimental Conditions and of Leisure Reading on the Subscales of the IRI-S D in Study 2*

| | **Emotional Concern** | | | **Perspective Taking** | | | **Fantasy** | | | **Personal Distress** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* |
| Intercept | 3.90 | 140.57 | **<.001** | 3.64 | 126.24 | **<.001** | 3.62 | 113.31 | **<.001** | 2.87 | 89.68 | **<.001** |
| Condition (priming) | 0.02 | 0.46 | .649 | 0.06 | 1.36 | .175 | -0.02 | -0.49 | .627 | 0.01 | 0.12 | .901 |
| Condition (non-priming) | 0.02 | 0.40 | .686 | 0.01 | 0.32 | .749 | 0.00 | 0.03 | .976 | -0.02 | -0.37 | .708 |
| ART | 0.16 | 0.77 | .441 | 0.44 | 1.99 | **.047** | 0.29 | 1.19 | .234 | -0.20 | -0.84 | .403 |
| Condition (priming) x ART | -0.07 | -0.24 | .814 | -0.46 | -1.51 | .132 | 0.02 | 0.07 | .942 | -0.01 | -0.03 | .973 |
| Condition (non-priming) x ART | -0.16 | -0.54 | .593 | 0.36 | 1.18 | .239 | 0.16 | 0.46 | .646 | -0.16 | -0.46 | .647 |
| Observations | | 408 | | | 408 | | | 408 | | | 408 | |
| $R^2$ / $R^2$ adjusted | | .00 / -.01 | | | .02 / .01 | | | .01 / -.01 | | | .00 / -.01 | |

*Note*. Categorical variables were effect coded (Condition: nonfiction = -1; priming = 1; non-priming = 1) and continuous variables were centered. ART = Author Recognition Test (Grolig et al., 2020; checklist B).

**Table A5**

*Effects of the Experimental Conditions Including the Post-Hoc No-Reading Control Group and of Leisure Reading on Theory of Mind and Empathy in Study 2*

| | **Theory of Mind** | | | | | | **Empathy** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* | *Estimates* | *t* | *p* |
| Intercept | 26.52 | 197.77 | **<.001** | 26.53 | 197.53 | **<.001** | 3.73 | 217.47 | **<.001** | 3.73 | 216.08 | **<.001** |
| Condition (priming) | -0.07 | -0.31 | .753 | -0.09 | -0.37 | .713 | 0.01 | 0.28 | .780 | 0.01 | 0.31 | .758 |
| Condition (non-priming) | -0.07 | -0.28 | .776 | -0.08 | -0.36 | .722 | 0.00 | 0.07 | .947 | -0.00 | -0.02 | .987 |
| Condition (nonfiction) | -0.05 | -0.22 | .824 | -0.11 | -0.48 | .628 | -0.03 | -1.06 | .288 | -0.04 | -1.22 | .222 |
| ART | | | | 3.04 | 3.03 | **.003** | | | | 0.28 | 2.14 | **.033** |
| Condition (priming) x ART | | | | 0.99 | 0.57 | .571 | | | | -0.15 | -0.67 | .503 |
| Condition (non-priming) x ART | | | | -1.74 | -1.00 | .316 | | | | 0.14 | 0.63 | .530 |
| Condition (nonfiction) | | | | 0.11 | 0.06 | .953 | | | | 0.07 | 0.30 | .765 |

| | | | | |
|---|---|---|---|---|
| Observations | 555 | 555 | 555 | 555 |
| $R^2$ / $R^2$ adjusted | .00 / -.00 | .02 / .01 | .00 / -.00 | .01 / -.00 |

*Note*. Categorical variables were effect coded (Condition: no-reading = -1; nonfiction = 1; priming = 1; non-priming = 1) and continuous variables were centered. ART = Author Recognition Test (Grolig et al., 2020; checklist B).