CrossMark

Validating self-paced sentence-by-sentence reading: story comprehension, recall, and narrative transportation

Ashley Chung-Fat-Yim[1] · Jordan B. Peterson[2] ·
Raymond A. Mar[1]

Abstract Previous studies on discourse have employed a self-paced sentence-by-sentence paradigm to present text and record reading times. However, presenting discourse this way does not mirror real-world reading conditions; for example, this paradigm prevents regressions to earlier portions of the text. The purpose of the present study is to investigate the ecological validity of self-paced sentence-by-sentence presentations by comparing it to normal page reading with respect to comprehension, recall, and narrative transportation, across two time points (immediate and delayed). Bayesian analyses found greater evidence in favor of the null hypothesis for transportation, indicating that little difference likely exists between sentence-by-sentence presentations and normal reading for this outcome. Weak evidence supporting the alternative hypothesis was found for immediate comprehension and recall, with participants who read the story as isolated sentences scoring marginally higher. Altogether, these results validate the self-paced sentence-by-sentence paradigm for measuring reading times, uncovering few differences in outcomes relative to natural reading.

Keywords Text presentation · Reading comprehension · Recall · Narrative transportation

✉ Ashley Chung-Fat-Yim
ashc88@yorku.ca

1   Department of Psychology, York University, Toronto, ON M3J 1P3, Canada

2   Department of Psychology, University of Toronto, Toronto, ON M5S 1A1, Canada

🍷 Springer

Introduction

In many studies that investigate discourse processes, reading time has been measured by presenting a text one sentence at a time, with participants asked to press a key to progress to the next sentence (e.g., Gerrig, Love, & McKoon, 2009; Rapp, 2008; Rapp & Taylor, 2004). This paradigm is similar in some ways to rapid serial visual presentation (RSVP; Forster, 1970), in which text is displayed word-by-word in the middle of the screen for set durations. It is also similar to the moving-window paradigm, in which readers press a key to reveal each successive word (Just, Carpenter, & Woolley, 1982). All three methods allow researchers to identify exactly when certain content is presented to participants and for how long. But reading a text in these ways departs from normal reading as it occurs in the real-world. With respect to the RSVP paradigm, saccadic eye movements typical of normal reading are eliminated entirely (Just & Carpenter, 1980), which results in changes in reading behavior such as an increase in attentional focus (Castelhano & Muter, 2001; Juola, Ward, & McNamara, 1982; Potter, 1984) and faster reading times (Öquist & Goldstein, 2002). Similarly, reading a narrative one sentence at a time prevents readers from backtracking or regressing to previous sentences, something that can occur quite often during everyday real-world reading (Rayner & Pollatsek, 1989; Vitu & McConkie, 2000). So although these paradigms provide a convenient way to either control or measure the flow of information during reading, it is important to consider whether adopting these approaches sacrifices external validity. It may be that reading a narrative one sentence at a time, contingent on a key-press, prevents readers from becoming fully engaged in the events being represented. Or, similarly, this method of text presentation might result in different reading strategies or different levels of effort while comprehending a text compared to typical reading. On the other hand, it may be that readers have little trouble adapting to a sentence-by-sentence presentation, such that they process the text and engage with its contents with ease. Given the prevalence of this paradigm in reading research, it is important to investigate the ecological validity of reading narratives in this way. In order to do so, we examined whether self-paced sentence-by-sentence presentations influence various reading outcomes, namely reading comprehension, text recall, and narrative transportation (i.e., feeling immersed and engaged with story events; Gerrig, 1993; Green & Brock, 2000).

To date, there is insufficient relevant research to make strong predictions as to whether sentence-by-sentence presentations are ecologically valid. Some have questioned this approach and similar methods, whereas other work seems to indicate that there is little issue with respect to ecological validity. For example, Schotter, Rayner, and Tran (2014) argue that systematically controlling eye movements in an RSVP paradigm decreases the accuracy of text comprehension, since this paradigm prevents the regressions that support higher-level linguistic processing (see also Baccino & Pynte, 1998). A faltering in comprehension typically triggers regressions, so preventing this referencing disallows readers from repairing their understanding. In their study, Schotter et al. (2014) asked participants to read garden-path sentences (i.e., sentences that are temporarily ambiguous, leading

readers toward a false interpretation) and unambiguous sentences, both normally and in a condition where the words became masked once participants moved their eyes. The authors found that preventing regressions within a sentence negatively impacted comprehension accuracy for both garden-path sentences and normal sentences. Since the sentence-by-sentence reading paradigm also prevents regressions, it is possible that this paradigm might have a similar influence on comprehension accuracy. A different study by Just et al. (1982) compared the self-paced moving-window paradigm to normal eye gaze with respect to word reading time and recall for short passages; these researchers found that the former produced longer reading times on each word but similar performance on recall. Based on the results from these two studies, presenting text in a way that prevents regressions appears to impair sentence comprehension, but preserve recall.

Previous studies have also examined differing ways of presenting discourse with the aim of uncovering the most preferred and efficient method of presenting text on small displays (Castelhano & Muter, 2001; Rahman & Muter, 1999). These studies employed short passages, often just one paragraph, and examined reading preferences, reading speed, and comprehension. The text was presented via self-paced sentence-presentation, self-paced page-presentation, or RSVP of single words presented at a predetermined rate. What these studies found was that sentence-presentation and page-presentation are equally preferred over RSVP, but no differences were observed between any of the text presentation formats with respect to comprehension (Castelhano & Muter, 2001; Rahman & Muter, 1999). Hence, these two studies show that sentence-by-sentence text presentations can be as efficient and effective as conventional page reading by some metrics, which is encouraging news for researchers who use the former to measure reading times. These studies, however, employed very short passages and so it would be valuable to investigate this question using longer, more ecologically-valid narratives, more akin to what is employed in discourse comprehension studies and how people read in the real-world. Doing so would allow us to determine the ecological validity of past work and also allow for more diverse aspects of reader response to be examined, such as engagement with the story.

Fictional narratives have been shown to impact readers in many ways. They can change our attitudes and beliefs (Appel & Mara, 2013; Escalas, 2004; Praxmarer, 2011; Prentice, Gerrig, & Bailis, 1997), self-concept (Djikic, Oatley, Zoeterman, & Peterson, 2009; Richter, Appel, & Calio, 2014), and possibly even our theory-of-mind abilities (Kidd & Castano, 2013; Koopman & Hakemulder, 2015; Mar & Oatley, 2008; cf. Panero et al., 2016). These outcomes seem related to a key feature of stories: enabling individuals to escape their everyday lives and enter a fictional world, forming strong emotional connections with characters and engaging with the plot. The experience of being immersed and absorbed into a fictional world is known as "transportation," based on the metaphor of being transported into the narrative world (Gerrig, 1993). Transportation is a distinct mental process integrating attention (or loss of access to real-world information), affect (emotional reactions), and ability to form mental imagery (Busselle & Bilandzic, 2009; Green & Brock, 2000). Stories that are well-crafted, have a high degree of realism, and report on events that are worth telling elicit greater transportation (Green, 2004).

Seeing as how transportation is an important aspect of engaging with fictional narratives, it is important to examine whether self-paced sentence-by-sentence reading has an impact on reader engagement. In order to do this, lengthier and more ecologically-valid texts must be presented, in comparison to the stimuli employed in past work. Thus, for our study we presented a professionally-written short story, *A Small, Good Thing* by Carver (1989). Employing a piece of professionally-written narrative fiction as a text not only allows us to study narrative transportation, but doing so also provides greater ecological validity compared to previous studies that employed only short paragraphs or sentences.

The present study thus aims to validate the self-paced sentence-by-sentence paradigm by investigating the effect of presentation format on comprehension, recall, and narrative transportation, whilst using a more realistic narrative text. Sentence-presentations were compared to self-paced presentation of whole pages, the latter of which better parallels a normal reading experience as it allows for regressions between and within sentences. If sentence-presentations result in approximately equivalent reading outcomes as page-presentations, this will provide an important validation of the ecological validity of this widely-used method of measuring reading times. In contrast, if sentence-presentations result in different outcomes, then past studies that rely on this paradigm will have to be examined with greater scrutiny with an eye toward determining if these results can be generalized to real-world reading.

Methods

Participants

One hundred and two students were recruited from an undergraduate research participant pool and all received academic credit for completing two separate testing sessions, held approximately 1 week apart. Eight participants were removed due to missing data. Hence, the final sample consisted of 94 participants (24 males, $M_{age} = 19.71$, $SD = 4.49$). Thirty-four out of 94 participants indicated they learned a language aside from English as their first language.

Measures and materials

Stimulus text

All participants were asked to read the entirety of a short story by Carver (1989), *A Small, Good Thing*. The story is about a young boy who dies 3 days after being struck by a car on the day of his birthday while walking to school. This story was chosen as it contains a great deal of emotional content, a suspenseful plotline, and is presented in an accessible and engaging prose style. It is 9635 words in length. This narrative was presented either one sentence at a time, controlled by the reader using a key-press (sentence-presentation), or one page at a time also self-paced (page-presentation), with the latter serving as a control condition that more closely mimics

real-world reading. Please see Table 1 for information on the level of difficulty of the text, assessed via the Lexile Framework for Reading (www.lexile.com) and Coh-Metrix Text Easability Assessor (http://tea.cohmetrix.com).

Immediate and delayed comprehension test

Comprehension of the text was measured using two separate tests comprising 10 multiple-choice questions (four response options per question), one presented directly after reading (Immediate) and the other a week later (Delayed). Items tested knowledge for specific details about the story, with some requiring the retrieval of factual details (e.g., "What age will Scotty be on his birthday?") and others requiring some degree of inference (e.g., "What made Ann realize who the caller was?"). All questions appear in the Supplementary Materials. Both scales had low levels of internal reliability, likely due to the variability in difficulty for each item along with the different types of story knowledge tested across items (Schmitt, 1996). The immediate comprehension test had a Cronbach's alpha of .16 and the delayed comprehension test had a Cronbach's alpha of .29.

Immediate and delayed free recall test

Memory for story details was measured by asking participants to recall as much information from the story as possible, both directly after reading (Immediate) and a week later (Delayed). Two individual coders coded the number of unique idea units recalled by each participant. For every unique detail of the story remembered, the participant was rewarded one point. The inter-rater reliability coefficients between two individual coders for immediate and delayed recall were .96 and .92, respectively.

Transportation scale

Engagement with the narrative was evaluated using the Transportation Scale developed by Green and Brock (2000). The scale contains 16 questions about participants' ability to visualize the events of the story, emotional engagement, and attentional focus. Responses were made using a 7-point Likert scale, ranging from "not at all" to "very much." Higher scores on the scale represent a greater degree of engagement with the text. Past research has shown that scores on this measure

Table 1 Text characteristics for "a small good thing" by raymond carver

Characteristics	Value
Average word length	4 letters
Average sentence length	10.22 words
Vocabulary level	3.1
Average log word frequency	3.66
Flesch Kincaid grade level	5.4

predict how persuaded people are by the themes of a narrative (van Laer, de Ruyter, Visconti, & Wetzels, 2014; Wyer, Adaval, & Colcombe, 2002). The scale had a Cronbach's alpha of .81 for the first session and .87 for the second session.

Demographic questionnaire

Participants were requested to provide demographic and background information, including gender, age, and years of education.

Procedure

In the first session, participants indicated their consent and then completed all measures on a computer. Participants were randomly assigned to one of two conditions: sentence-presentation ($n = 47$) or page-presentation ($n = 47$). They then read the target story either one sentence at a time or one page at a time, hitting the spacebar or clicking the mouse to progress in both cases. Text was displayed in black 14-point Times New Roman font on a white background (600px × 800px, resolution). For sentence-presentation, sentences were centered on the screen horizontally, consistent with common research methodology (Rahman & Muter, 1999). For page-presentation, text was single-spaced, left justified with paragraph indentations included to be consistent with real-world text presentations. Participants were allowed to re-read within the sentence or page; however once they pressed the spacebar they were unable to go back to the previous page or sentence. The order of presentation for the other measures was the same for all participants. Once the reading of the narrative was completed, participants were administered the Transportation Scale, Immediate Free Recall Test, and Immediate Comprehension Test. A general demographics questionnaire was administered last. The second session took place approximately 1 week later, during which participants were administered the Delayed Free Recall Test, Delayed Comprehension Test, and Transportation Scale once again.[1] At the end of the second session, participants were debriefed by the experimenter and received partial course credit for their time. Each testing session took approximately 1 h to complete.

---

[1] Participants also completed additional measures not discussed here, namely a measure of print-exposure (Author Recognition Test; Stanovich & West, 1989), general reading ability (Component Processes Task; Hannon & Daneman, 2001), and Five Factor personality traits (Big Five Personality Inventory, BFI; John & Srivastava, 1999). These measures did not seem to play a large role in our results and we lack the statistical power to conduct reliable moderation analyses. Nevertheless, exploratory moderation analyses found that reading ability and print-exposure did not tend to moderate associations between condition and comprehension, recall, or transportation at either time point. Additionally, the personality traits conscientiousness, extraversion and emotional stability did not moderate associations between condition and comprehension, recall, or transportation at either time point. However, there was a significant difference between text presentation and recall for those high in agreeableness at both time points, with sentence-presentation being associated with better recall [Immediate recall: $t(45) = 2.50$, $p = .016$, $d = .73$; Delayed recall: $t(45) = 3.11$, $p = .003$, $d = .90$], but not for those low in agreeableness, all $ps > .21$. There was also a significant difference between conditions on delayed comprehension for those high in openness, $t(45) = 2.36$, $p = .023$, $d = .70$, but not for those low in openness, $p = .63$. The supplementary materials include the full results of these moderation analyses.

Results

Mean scores and standard deviations by text presentation format and session for all measures are reported in Table 2. Participants in the sentence-presentation and page-presentation conditions were roughly equivalent in age and years of education, Age: $t(92) = .21$, $p = .84$, $d = .04$, 95 % CI [−.45, .36]; Education: $t(92) = .38$, $p = .70$, $d = .08$, 95 % CI [−.48, .32].

Bayesian independent samples t tests were performed to examine the likelihood that the two presentation formats result in differences in comprehension, recall, and transportation (H1), compared to the likelihood that the two presentations result in equivalent outcomes (H0), using JASP v.0.7.5.5 (JASP Team, 2016; jasp-stats.org). Bayesian model selection offers the possibility to use prior knowledge to compare hypotheses based on specific expectations. The output is a Bayes Factor (BF), which provides information on the probability or likelihood of the null and alternative hypothesis (Kass & Raftery, 1995; Rouder, Speckman, Sun, Morey, & Iverson, 2009). A BF_{01} is the likelihood of the data given the null hypothesis over the likelihood of the data given the alternative hypothesis (BF_{10} is the opposite). For example, a BF_{01} of 2.00 would indicate that the data observed are twice as likely under the null hypothesis (of no difference) compared to the alternative hypothesis (of a difference, the magnitude and range of which is specified by the prior). One advantage of using Bayesian statistics is that the assumptions are not based on normality, making it a suitable approach for relatively small sample sizes (Gill, 2014). Moreover, Bayesian analyses allow researchers to evaluate the likelihood of the null hypothesis, unlike traditional null hypothesis statistical testing (NHST), which assumes the null is true (Fraley & Marks, 2007).

Table 2 Mean scores and standard deviations on all measures by text presentation format and session

	Sentence M (SD)	Page M (SD)
Background information		
Age in years	19.62 (4.22)	19.81 (4.79)
Years of education	13.38 (2.10)	13.62 (3.64)
English as L1 (%)	68.09 %	59.57 %
Overall reading time	42.91 min (12.00)	37.23 min (9.55)
Session 1		
Immediate free recall test	67.81 (15.62)	62.27 (15.68)
Immediate comprehension test	8.83 (0.99)	8.38 (1.24)
Transportation scale	4.74 (0.86)	4.54 (0.79)
Transportation vivid imagery subscale	5.11 (1.08)	4.77 (1.15)
Session 2		
Delayed free recall test	52.18 (14.69)	48.38 (12.89)
Delayed comprehension test	7.53 (1.49)	7.13 (1.45)
Transportation scale	4.48 (0.85)	4.36 (0.92)
Transportation vivid imagery subscale	4.83 (1.15)	4.61 (1.25)

Results for the immediate and delayed free recall test is reported in terms of idea units

Bayesian independent *t* tests were performed using the Jeffrey–Zellner–Siow Prior (Jeffreys, 1961), which is a prior used in a previous reading study (Abbott & Staub, 2015). In addition, a cauchy prior width of .40 was chosen because it corresponds to a medium effect size (Cohen, 1988). See Fig. 1 for the Bayes factor
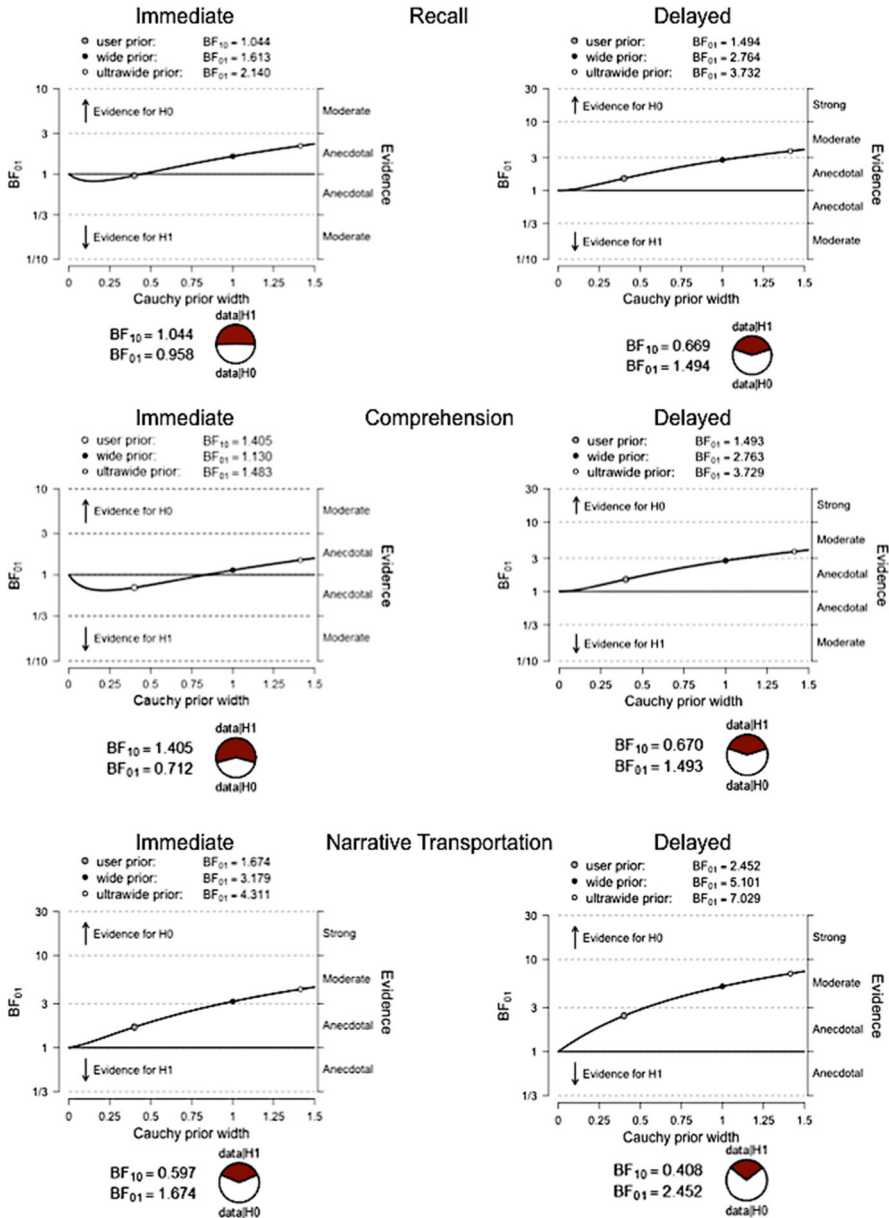

Fig. 1 Bayes factor robustness check for immediate and delayed comprehension, recall, and narrative transportation. All figures are taken directly from the JASP output

robustness check of each measure within each session. These figures illustrate how the results would change if wider priors, indicating greater a priori uncertainty regarding the likelihood of an effect, were adopted. In light of the dearth of past research, greater uncertainty regarding the outcome would seem to be warranted.

We found that overall reading times were longer for the sentence-presentation condition than page-presentation condition, $t(92) = 2.54$, $p = .013$, $d = .52$, 95 % CI [.11, .93], $BF_{01} = .25$, $BF_{10} = 3.97$. For the measures administered in the first session, participants were about equally engaged with the story across conditions, as measured by the Transportation Scale, $t(92) = 1.21$, $p = .23$, $d = .25$, 95 % CI [$-.17$, .65], $BF_{01} = 1.67$, $BF_{10} = .60$.[2] And so, the data are roughly 1.7 times more consistent with there being no difference in narrative engagement based on whether one is reading sentence-by-sentence or a page at a time. Small differences were observed for both immediate recall and comprehension, with those in the sentence-presentation condition exhibiting better performance in both cases. The NHST results fell just above traditional threshold for statistical significance (and with confidence intervals that include zero) and the Bayesian results indicated weak evidence in favor of H1 for immediate comprehension and immediate recall, Immediate Comprehension Test: $t(92) = 1.93$, $p = .057$, $d = .40$, 95 % CI [$-.01$, .81], $BF_{01} = .71$, $BF_{10} = 1.41$; Immediate Free Recall Test: $t(92) = 1.72$, $p = .089$, $d = .35$, 95 % CI [$-.06$, .76], $BF_{01} = .96$, $BF_{10} = 1.04$. Comprehension and recall were weakly, and not statistically significantly, related immediately after reading the short story, $r(94) = .16$, $p = .13$, 95 % CI [$-.05$, .35].

For measures administered in the second session, participants in the sentence-presentation condition did not differ much from those in the page-presentation condition for our measures of narrative engagement, memory, or recall. In all cases the Bayesian result found either greater evidence in favor of the null, or no evidence in favor of either the null or the alternative hypothesis, Transportation Scale: $t(92) = .68$, $p = .50$, $d = .14$, 95 % CI [$-.27$, .54], $BF_{01} = 2.45$, $BF_{10} = .41$; Delayed Free Recall Test: $t(92) = 1.33$, $p = .19$, $d = .27$, 95 % CI [$-.13$, .68], $BF_{01} = 1.49$, $BF_{10} = .67$; Delayed Comprehension Test: $t(92) = 1.33$, $p = .19$, $d = .27$, 95 % CI [$-.14$, .68], $BF_{01} = 1.49$, $BF_{10} = .67$. Comprehension and recall were moderately associated when measured during this second session, $r(94) = .37$, $p < .001$, 95 % CI [.18, .53]. See Fig. 1 for Bayes factor robustness check for each measure and session.

Discussion

The goal of the present study was to investigate the ecological validity of implementing a sentence-by-sentence paradigm when examining the reading of discourse, using a longer and more ecologically-valid text compared to past studies.

---

[2] A Vivid Imagery subscale was created based on the items of the Transportation Scale because there were seven items that focused on one's ability to visualize the narrative. These items were 1, 3, 12, 13, 14, 15 and 16. This subscale had a Cronbach's alpha of .84. Participants in the sentence-presentation condition did not differ much in their ability to visualize the narrative from those in the page-presentation condition in both sessions, Session 1; $t(92) = 1.48$, $p = .14$, $d = .31$, 95 % CI [$-.82$, .00], $BF_{01} = 1.28$, $BF_{10} = .78$; Session 2: $t(92) = .90$, $p = .37$, $d = .19$, 95 % CI [-1.71, $-.82$], $BF_{01} = 2.15$, $BF_{10} = .47$.

To achieve this, a short story high in emotional content was presented as sentences and the outcomes were compared to those observed after reading the same story presented as individual pages. In general, we found little evidence that reading a story sentence-by-sentence leads to differences in long-term comprehension and recall, compared to reading one page at a time. A major goal of this study was to examine whether different text presentation paradigms influence how engaged readers become with a story. We observed no differences between text presentation formats with respect to narrative transportation, validating past work that relies on sentence-by-sentence presentation paradigms with respect to narrative engagement. Overall, these findings provide an important validation of the sentence-by-sentence paradigm, indicating that it can be employed to study reading times with little fear that reading processes and outcomes will not resemble those observed during real-world naturalistic reading.

On the surface, self-paced sentence-presentation formats look like a rather artificial method of reading, raising the question of whether this approach provides an experience that is ecologically valid. However, participants were equally engaged in the storyline across the two forms of text presentation, demonstrating that sentence-by-sentence paradigms also have the ability to absorb readers into a fictional world, where they identify with characters and feel emotionally involved with the narrative. In addition, the differences in comprehension and recall across presentation formats appear to be rather minimal.

Some exceptions to this lack of difference between conditions were observed. In particular, participants read the short story slower when it was presented as isolated sentences compared to pages. As a result, reading time measurements based on sentence-by-sentence presentations may be different from more those found during natural reading based on pages of text. That said, in most cases researchers are interested in differences in reading times between key sentences rather than the reading times themselves and our data do not permit any inferences regarding whether differences in reading time are affected by sentence-presentations. The sentence-presentation condition also resulted in higher scores for comprehension and recall when measured immediately after reading the short story. Although the magnitude of these differences was not large, Bayesian analyses found greater evidence for the alternative hypothesis (a difference between presentations) for immediate comprehension, recall, and overall reading times. One possible explanation for these differences is that readers in the sentence-presentation condition may have been more likely to read each sentence carefully, knowing that this was the only opportunity to process this information (i.e., it would be impossible to re-read the sentence by regressing to it one a key has been pressed). In contrast, readers in the page-presentation condition may have read more quickly, skimming the text because they felt overly comfortable in the knowledge that they could re-read parts of the page if they felt they did not understand what was happening. Hence, it is possible that readers of each presentation type employed different strategies. These are purely speculative post hoc interpretations, however, and it should be emphasized that the differences between conditions for these outcomes were not large.

One limitation of the current study is that only a single narrative was presented and therefore it is unclear whether the results generalize to other short stories or

lengthier pieces of published fiction. It is possible that with longer pieces of fiction, the small effects in favor of sentence-presentation for immediate comprehension will be diluted and possibly disappear. In addition, it is worth noting that Raymond Carver's style is straightforward and rather expository, which may have reduced the readers' need to regress to earlier portions of the text. Other styles of writing (e.g., frequent use of free indirect discourse) may be more challenging and more likely to necessitate regressions, resulting in poorer comprehension for sentence-presentations. Hence, future studies should aim to examine the ecological validity of the sentence-by-sentence presentation paradigm with longer narratives and stories that contain other styles of writing (e.g., word play, and allusions). It would also be useful to replicate this study using scales that capture other relevant experiences associated with reading that are not captured by the transportation scale (e.g., reflection and insight; Miall & Kuiken, 1995). Furthermore, this study could be improved upon in the future by employing a larger sample. This would improve statistical power and reduce the width of confidence intervals around our point estimates. The Bayesian analyses we report are independent of sample size, however, and so are not influenced by this factor.

In conclusion, our data demonstrate that it is appropriate to use sentence-reading paradigms in discourse research. Based on these findings, any differences in how readers process a text sentence-by-sentence versus page-by-page do not appear to manifest large differences for outcomes like comprehension, recall, or engagement. As a result of this study, past research on narrative that utilizes a sentence-presentation paradigm can now be viewed with greater confidence with respect to ecological validity.

References

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z Reader's null predictions. *Journal of Memory and Language, 85*, 76–87. doi:10.1016/j.jml.2015.07.002.

Appel, M., & Mara, M. (2013). The persuasive influence of a fictional character's trustworthiness. *Journal of Communication, 63*(5), 912–932. doi:10.1111/jcom.12053.

Baccino, T., & Pynte, J. (1998). Spatial encoding and referential processing during reading. *European Psychologist, 3*(1), 51–61. doi:10.1027/1016-9040.3.1.51.

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology, 12*(4), 321–347. doi:10.1080/15213260903287259.

Carver, R. (1989). A small, good thing. In *Where I'm calling from: Selected stories*. New York: Vintage.

Castelhano, M. S., & Muter, P. (2001). Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour and Information Technology, 20*(4), 237–247. doi:10.1080/01449290110069400.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Djikic, M., Oatley, K., Zoeterman, S., & Peterson, J. B. (2009). On being moved by art: How reading fiction transforms the self. *Creativity Research Journal, 21*(1), 24–29. doi:10.1080/10400410802633392.

Escalas, J. E. (2004). Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising, 33*(2), 37–48. doi:10.1080/00913367.2004.10639163.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception and Psychophysics, 8*(4), 215–221. doi:10.3758/BF03210208.

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance testing debate and its implications for personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 149–169). New York: Guilford.

Gerrig, R. J. (1993). *Experiencing narrative worlds*. New Haven, CT: Yale University Press.

Gerrig, R. J., Love, J., & McKoon, G. (2009). Waiting for Brandon: How readers respond to small mysteries. *Journal of Memory and Language, 60*(1), 144–153. doi:10.1016/j.jml.2008.08.004.

Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach* (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Green, M. C. (2004). Transportation into the narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes, 38*(2), 247–266. doi:10.1207/s15326950dp3802_5.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*(5), 701–721. doi:10.1037/0022-3514.79.5.701.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*(1), 103–128. doi:10.1037/0022-0663.93.1.103.

JASP Team. (2016). JASP (Version 0.7.5.5) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York: Oxford University Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 102–138). New York: Guilford Press.

Juola, J. F., Ward, N. J., & McNamara, T. (1982). Visual search and reading of rapid serial presentations of letter strings, words, and text. *Journal of Experimental Psychology: General, 111*(2), 208–227. doi:10.1037/0096-3445.111.2.208.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. doi:10.1037/0033-295X.87.4.329.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228–238. doi:10.1037/0096-3445.111.2.228.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. doi:10.2307/2291091.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*(6156), 377–380. doi:10.1126/science.1239918.

Koopman, E., & Hakemulder, F. (2015). Effects of literature on empathy and self-reflection: A theoretical-empirical framework. *Journal of Literary Theory, 9*(1), 79–111. doi:10.1515/jlt-2015-0005.

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

Miall, D. S., & Kuiken, D. (1995). Aspects of literary response: A new questionnaire. *Research in the Teaching of English, 29*, 37–58.

Öquist, G., & Goldstein, M. (2002). Towards an improved readability on mobile devices: Evaluating adaptive rapid serial visual presentation. *Interacting with Computers, 15*(4), 539–558. doi:10.1016/S0953-5438(03)00039-0.

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

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (pp. 91–118). Hillsdale, NJ: Erlbaum.

Praxmarer, S. (2011). Message strength and persuasion when consumers imagine product usage. *Journal of Consumer Behaviour, 10*(4), 225–231. doi:10.1002/cb.331.

Prentice, D. A., Gerrig, R. J., & Bailis, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review, 4*(3), 416–420. doi:10.3758/BF03210803.

Rahman, T., & Muter, P. (1999). Designing an interface to optimize reading with small display windows. *Human Factors, 41*(1), 106–117. doi:10.3758/BF03210803.

Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory and Cognition, 36*(3), 688–710. doi:10.3758/MC.36.3.688.

Rapp, D. N., & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30*(5), 988–1001. doi:10.1037/0278-7393.30.5.988.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Richter, T., Appel, M., & Calio, F. (2014). Stories can influence the self-concept. *Social Influence, 9*(3), 172–188. doi:10.1080/15534510.2013.799099.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. doi:10.3758/PBR.16.2.225.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353. doi:10.1037/1040-3590.8.4.350.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science, 25*(6), 1218–1226. doi:10.1177/0956797614531148.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*(4), 402–433. doi:10.2307/747605.

van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2014). The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research, 40*(5), 797–817. doi:10.1086/673383.

Vitu, F. M., & McConkie, G. W. (2000). Regressive saccades and word perception in adult perception. In A. Kennedy, R. Radach, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 89–118). Amsterdam: Elsevier.

Wyer, R. S., Adaval, R., & Colombe, S. J. (2002). Narrative-based representations of social knowledge: Their construction and use in comprehension, memory, and judgment. In M. P. Zanna (Ed.), *Advanced in experimental social psychology* (Vol. 34, pp. 131–197). New York: Academic Press.